Statistics of Gradient Directions in Natural Images

Alexandre James Nasrallah

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of the

University of London.

Department of Computer Science

University College London

August 19, 2007

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

I, Alexandre James Nasrallah, declare that the work presented in this thesis is my own. Where the work is not my own, it has been stated.

This dissertation has not been presented to any other University for examination in either the United Kingdom or Overseas.

Signed:

Date: 19 August 2007

Abstract

Interest in finding statistical regularities in natural images has been growing since the advent of information theory and the advancement of the efficient coding hypothesis that the human visual system is optimised to encode natural visual stimuli. In this thesis, a statistical analysis of gradient directions in an ensemble of natural images is reported. Information-theoretic measures have been used to compute the amount of dependency which exists between triples of gradient directions at separate image locations. Control experiments are performed on other image classes: phase randomized natural images, whitened natural images, and Gaussian noise images.

The main results show that for an ensemble of natural images the average amount of dependency between two and three gradient directions is the same as for an ensemble of phase randomized natural images. This result does not extend to i) the amount dependency between gradient magnitudes, ii) gradient directions at high gradient magnitude locations, or iii) individual natural images. Furthermore, no significant synergetic dependencies are found between triples of gradient directions in an ensemble natural images; a synergetic dependency is an increase in dependency between a pair of gradient directions given the interaction of a third gradient direction.

Additional experiments are performed to establish both the generality and specificity of the main results by studying the gradient direction dependencies of ensembles of noise (random phases) images with varying power law power spectra. The results of the additional experiments indicate that, for ensembles of images with varying power law power spectra, the amount of dependency between two and three gradient directions is determined by the ensemble's mean power spectrum rather than the phase spectrum. A framework is also presented for future work and preliminary results are provided for the dependency between second order derivative measurements (shape index) for up to 9-point configurations.

Acknowledgements

Although there are other people I could mention, I would like to reserve acknowledgement to the few who made a significantly larger impact on me whilst working on this thesis.

First, I would like to thank my supervisor, Lewis D Griffin, for his encouragement throughout, positive attitude, always being available for a discussion (I cannot remember a single occasion when he was not available!), and for guiding me towards a wide variety of relevant talks—all of which helped to give me a fuller appreciation of the subject. I would also like to thank him for his trust in me during difficult times.

Second, I would like to thank my dear Sasha, for providing meaning to life when I could find little meaning in my results. I have been fortunate to keep her company in what can be a lonely endeavour and a trying one for others to be around.

Third, I thank my parents for their support which has afforded me a youth devoted to seeking knowledge and understanding. Without the environment they provided, I could not imagine having the opportunity to study what I love.

Contents

1 Introduction **10**

1.1 Preview of Thesis 11

1.1.1 Natural Images 11

1.1.2 Visual Perception: Gestalt Theory 11

1.1.3 Multi-local Statistics 12

1.1.4 Information Theory 12

1.1.5 Gradient Directions 14

1.1.6 Higher Order Statistics 14

1.1.7 Summary of Motivation 16

1.2 Efficient Coding Hypothesis 17

1.3 Overview of Thesis 20

2 Background **22**

2.1 Light 22

2.1.1 Properties of Light 23

2.1.2 Inverse Optics 25

2.2 Human Visual System 28

2.2.1 Retina 29

2.2.2 Primary Visual Cortex 33

2.3 Human Visual Perception 37

2.3.1 Unconscious Inference and Likelihood Principle 38

2.3.2 Gestaltists 40

2.3.3 Gestalt Principles of Grouping 41

2.3.4 Global Versus Local Approach to Perception 43

2.4 Computational Approach to Vision 45

2.4.1 Image Features . . . 45
2.4.2 Derivative Operators as Feature Detectors . . . 46
2.4.3 Gauge Coordinates and Gradient Direction . . . 48
2.4.4 Corner Detectors . . . 48
2.4.5 Second Order Gauge Frame and Shape Index . . . 50

3 Natural Image Statistics **54**

3.1 Introduction . . . 54
3.2 Background: Fourier Analysis . . . 55
3.2.1 Local Energy . . . 58
3.3 Power Spectrum . . . 59
3.3.1 Power Law Scaling and Scale Invariance . . . 60
3.3.2 Sources of Scale Invariance . . . 61
3.3.3 Variation in Natural Image Power Spectra . . . 62
3.3.4 Psychophysical and Physiological Significance of Power Spectra . . . 62
3.4 Contrast Statistics . . . 64
3.4.1 Contrast Adaptation . . . 65
3.4.2 Non-Gaussian Statistics of Natural Images . . . 66
3.5 Higher Order Statistics . . . 67
3.5.1 Gabor Filters . . . 67
3.5.2 Phase Spectra . . . 70
3.5.3 Psychophysical Significance of Phase Spectra . . . 71
3.5.4 Geometric Properties of Natural Images . . . 72
3.6 Summary of Natural Image Statistics . . . 76

4 Technical Aspects to Computations and Methodology **78**

4.1 Information Theory . . . 79
4.1.1 Shannon's Measure of Entropy . . . 79
4.1.2 Mutual Information . . . 81
4.1.3 Interaction Information: Synergy and Redundancy . . . 81
4.2 Entropy Estimators . . . 85
4.3 Variance of Entropy Estimators: Bootstrap . . . 86

4.4 Applying Gaussian Derivatives to Images 89
4.4.1 Linear Scale-Space Theory 90
4.4.2 Errors Gaussian Derivative Operators 91

5 Methodology: Data Collection **93**
5.1 Generating Image Classes 93
5.2 Extracting Gradient Directions and Magnitude 95
5.3 Data Collection 96
5.4 Overlap of Kernels 100
5.5 Entropy Estimates from Histogram Data 100
5.6 Bootstraps 101

6 Results **103**
6.1 1-point Statistics 103
6.2 2-point Statistics 108
6.2.1 Gradient Direction Dependencies 109
6.2.2 Gradient Magnitude Dependencies 109
6.2.3 Strong Gradient Direction Dependencies 111
6.3 Gradient Direction Dependencies of Individual Natural Images 112
6.4 3-point Statistics 114
6.4.1 Gradient Direction Dependencies 116
6.4.2 Gradient Magnitude Dependencies 118
6.4.3 Strong Gradient Direction Dependencies 119
6.5 Discussion of Main Results: 1-, 2-, 3-point Statistics 119

7 Further Experiments **123**
7.1 Analysis of Other Synthetic Images 123
7.1.1 Varying Power Spectral Slope Images 124
7.1.2 Dead Leaves Model Images 130
7.1.3 Constant-Sized Disk Images 132
7.1.4 Conclusion of Further Experiments on Synthetic Images 134
7.2 Information Contour Map of 2-point Statistics 134
7.2.1 Methodology: Information Contour Maps 135

7.2.2 Results: Information Contour Maps 137
7.2.3 Conclusions: Information Contour Maps 142
7.3 Experiments to find the Source of Synergetic Dependencies 142
7.3.1 Line Configuration 142
7.3.2 Results and Discussion: Line Configuration 143
7.3.3 Curl of a Gradient Field 143
7.3.4 Results and Discussion: Curl of Gradient Field 145
7.3.5 Other Random Distributions 147
7.4 Gradient Direction Configurations 147
7.5 Conclusions of Further Experiments 150

8 Preliminary Future Work and Conclusions 152
8.1 Preliminary Future Work 152
8.1.1 Extending Information Theory to 9-point Dependencies 154
8.1.2 Methodology for 9-point Analysis 156
8.1.3 Preliminary Results 157
8.1.4 Conclusions to Preliminary Work 159
8.2 Conclusions 163
8.2.1 Summary of Background and Motivation 163
8.2.2 Summary of Relevant Work 164
8.2.3 Summary of Methodology 164
8.2.4 Main Findings 165
8.2.5 Limitations of Work 168
8.2.6 Final Remarks 170

Appendices 173

A Properties of Light 173
A.1 Depth Cues 173

B Human Visual System 175
B.1 Neuron Signals 175
B.2 Photoreceptor Cells 176

B.3 Optic Chiasm and Lateral Geniculate Nucleus 177

B.4 Hypercolumns 179

C Human Visual Perception **181**

C.1 Philosophical Implications of Gestalt Theory 181

C.2 Holism 182

C.3 Memory and Holistic Perception 183

C.4 Common fate and Closure 184

C.5 Multistability 185

C.6 Description of a Multistable Neuron Network 186

C.7 Competition between Gestalt Grouping Principles 187

C.8 Quantifying Figural Goodness 188

D Computational Approach to Vision **190**

D.1 Edge Detection Algorithms 190

Bibliography **193**

List of Publications **210**

Chapter 1

Introduction

Vision science is a broad subject which incorporates a number of disciplines: physics (optics), mathematics (statistics, geometry), engineering (signal processing), biology (neuroscience) and psychology (cognitive psychology). An interdisciplinary approach is necessary if any thorough understanding of visual systems—in particular the human visual system (HVS)—is to be achieved. In any mammalian species, the visual system provides an important representation of the physical world and is a vital component in the survival of the species.

In the case of the HVS, advances towards an improved understanding of this system have been inspired by the assumption that the HVS is optimised to encode natural stimuli, which is known as the efficient coding hypothesis proposed by Attneave [1]. The hypothesis posits that the HVS is specifically adapted to signals of the natural environment because these are the most common signals the HVS receives. In the context of neuron responses in the HVS, Barlow hypothesised that the role of such neurons is to remove the statistical redundancies found in natural signals [2, 3]. The advent of information theory and the efficient coding hypothesis has led to a profusion of research to find the statistical regularities in the natural environment.

Other factors are likely to have influenced how the HVS has evolved to its current state. For example, any biological visual system must enable the species to perform essential tasks which are necessary for its survival. The hardware of the system will limit how much and how quickly information can be transferred within the system whilst remaining energetically sustainable. The hardware must also be robust to the environment: a visual system must—within reason—remain undamaged by stimuli. For example, it is known that the HVS is only sensitive to a certain range of wavelengths in the electromagnetic spectrum. The range of sensitivity is, approximately, from 400nm to 700nm. This range does not include infrared and ultraviolet radiation, which, if absorbed by the visual system, would cause undesirable physical changes

to its living cells. Nevertheless, it is possible to find living species which are sensitive to infrared (pit viper snake) and ultraviolet light (bees and birds). However, for these species, the advantages outweigh the costs of damage to its cells. The evolutionary costs of the hardware is another important factor in its evolution because for greater complexity in the hardware, the greater the risk and cost associated with encoding it in the DNA of the species.

1.1 Preview of Thesis

In the next section, several topics, which are covered in more depth in later sections of this thesis, will be introduced here as motivation for the rest of this work.

1.1.1 Natural Images

Whatever evolutionary factors may have determined the state of the HVS today, light from the environment is projected (upside down) onto a retina at the back of the eye. This results in a 2-D image being formed from a 3-D world in the HVS, and therefore the HVS must be performing all of its processing on 2-D images. Thus, in this thesis, it is assumed that an ensemble of natural images (any image taken from the natural environment) is an adequate representation of the natural environment. Moreover, this thesis is concerned with low-level image analysis and, therefore, colour and motion (i.e. video or series of images) are ignored. A study, by Caselles et al., showed that for a natural image there is no geometric information contained in the colour channels of an image which is not present in its grey level image [4].

There has been a considerable amount of work on finding methods to extract low-level features from 2-D images. In particular, several low-level image features are extracted by using polynomial combinations of derivative operators. This computational approach to vision has been aided by Marr's paradigm that visual perceptual tasks can be formulated in terms of information processing tasks and this approach is presented in Section 2.4.

1.1.2 Visual Perception: Gestalt Theory

Despite Marr's paradigm, a 2-D image remains flat. Thus, it is likely that in processing 2-D images the HVS is making additional assumptions in order to construct a 3-D world. The problem of the 'missing dimension' in the HVS is known as the problem of 'inverse optics', which is strongly linked to theories of visual perception. Gestalt theory of human perception, in particular perceptual grouping, along with other vision theories (e.g. structuralism, unconscious

inference) are presented in Section 2.3.

Gestalt psychologists were able to establish a number of qualitative grouping rules which govern human visual percepts. A motive for the methodology adopted in this thesis is an attempt at relating the Gestalt philosophy of holism to the statistics of natural images by finding quantitative evidence for the advantages—in an information-theoretic sense—in making multi-local geometrically meaningful (e.g. gradient directions) measurements in images, whereby multi-local is defined to mean local measurements made simultaneously at multiple locations in an image.

1.1.3 Multi-local Statistics

In this work, multi-local statistics are classified firstly by the derivative order of the image measurements considered, and secondly, by the number of locations at which these measurements are made. For instance, the histogram of the distribution of gradient directions, resulting from randomly collecting gradient directions from an ensemble of natural images, is termed 1-point first order statistics (see Fig. 1.1 for an illustration of gradient orientations in a natural image). When the mutual information between two gradient directions at separate locations is computed, the result is classified as 2-point first order statistics. This is then extended to measure the dependencies of gradient directions at three separate locations which is classified as 3-point first order statistics.

1.1.4 Information Theory

To quantify the amount of dependency between gradient directions in natural images information-theoretic methods have been used, specifically interaction information (higher order mutual information). For example, triples of gradient directions are measured and the mutual information between a pair of gradient directions is computed. Furthermore, how the context of a third gradient direction affects the 2-point mutual information is calculated. If, for example, 2-point dependency increases in the context of a third point, then measuring triples of gradient directions is termed synergetic. This means there is an additive (i.e. 'holistic') effect in measuring triples of gradient directions simultaneously, however, if there is a decrease, the 3-point dependency is termed redundant. If there is neither an increase nor decrease, then there is no triple interaction.

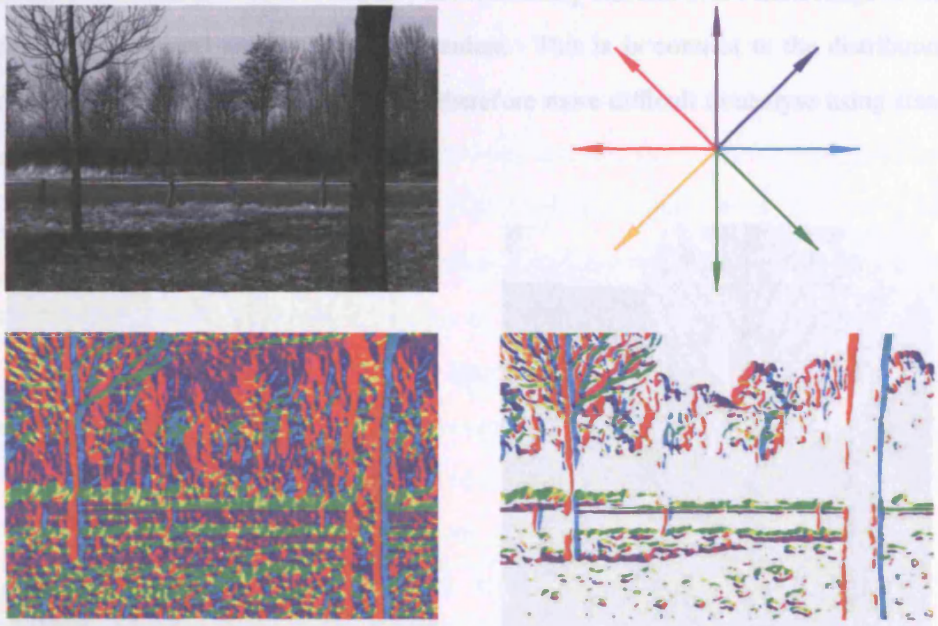


Figure 1.1: Top left: an example of a natural image. Top right: colour wheel key of gradient directions, e.g. red and cyan denote horizontally oriented gradient directions. Bottom left: the natural image above has been converted to a colour map of gradient directions. Bottom right: same as bottom left, but with only the strong (top 25% gradient magnitudes) gradient directions shown; applying this threshold to the gradient directions makes it resemble an edge orientation map.

1.1.5 Gradient Directions

In order to search for multi-local dependencies in natural images, there are several reasons why it is appropriate and interesting to analyse gradient directions. One reason is because gradient directions describe the extrinsic geometry of a location in the image and are invariant to monotonic transformations (global brightness and contrast changes) of the image, as shown in Figure 1.2. Thus, the gradient direction together with the gradient perpendicular form a natural local coordinate frame (first order gauge), which is discussed in Section 2.4.3. A further advantage of gradient directions is that they are reasonably uniform over a finite range of values, which makes statistical analysis more convenient. This is in contrast to the distribution of filter responses which can be kurtosed and therefore more difficult to analyse using standard statistical methods.

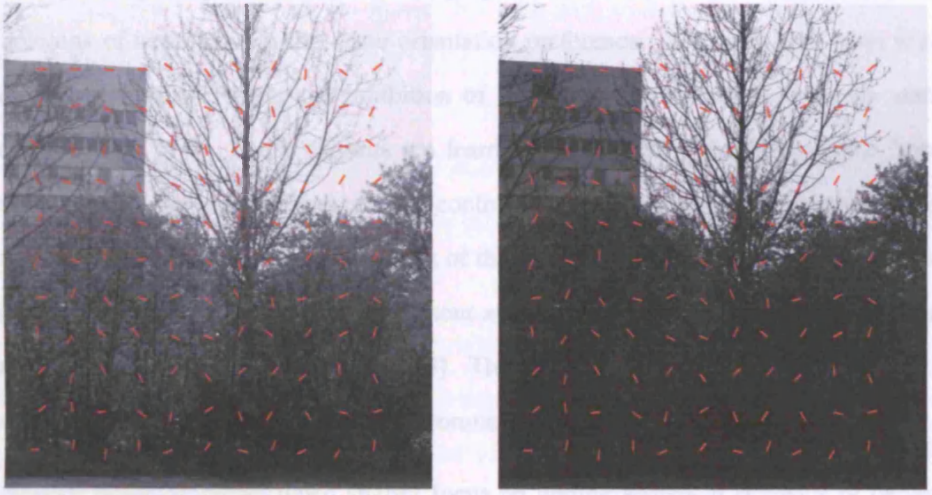


Figure 1.2: Left side: illustrates a natural image with red lines oriented normal to the gradient direction. Right side: illustrates the same natural image but at a lower contrast, notice how the gradient norms remain in the same direction despite the global change in contrast.

1.1.6 Higher Order Statistics

The choice of gradient directions is further motivated by the discovery, by previous authors, of higher order statistical regularities in ensembles of natural images (a detailed discussion is presented in Chapter 3.5). For example, a prevalence of horizontally and vertically oriented contours [5, 6, 7]. Other authors have also observed the responses of edge detector-like filters and found that greater spatial dependencies exist between pairs of edges in natural images

compared to synthetic images (power spectrum of natural images but with randomized phases). Moreover, in natural images, the dependency between edges is greatest when the second edge is along the initial edge direction [8]. This implies that the dependencies between edges in natural images are affected by their relative orientation and position to each other. Studies also reveal that statistically the most likely contour joining two nearby edge segments is a line and then a circular arc [9]. Therefore, initially it was proposed in [9] that natural images have an abundance of circles, however, further investigations in [10] reveal that natural scenes are in fact likely to have many closed smooth contours, not just circles.

The observations of the co-occurrence statistics of edges in natural images provides some evidence for why cortical neurons in the primary visual area (V1) might be sensitive to lines of particular orientation (a discussion of V1 physiology is in Section 2.2.2). It has been found that in the cat [11, 12] and primate visual cortex [13] there exist lateral connections which mostly link columns of neurons with the same orientation preference. Attempts have been made to learn the pattern of excitation and inhibition of these lateral connections from the statistics of natural images [14]. These patterns are learned by integrating the information from the responses of spatially-nearby filters to build contours [15]. The orientation selectivity found in neurons has been hypothesised to be a result of the natural environment. However, analysis of test images full of oriented features but without any second order correlations do not result in oriented receptive fields being predicted [16]. Therefore, orientation selectivity may not be a consequence of oriented features in the environment.

Several of the aforementioned studies focus on finding statistical evidence of dependencies between pairs of edges and contours, both of which can act as cues to identify features in an image. However, edges are only found at locations in the image with high gradient magnitude. Furthermore, apart from a previous study of 2-pixel and 3-pixel luminance correlations in natural images [17], which revealed that the contribution of 3-pixel correlations is extremely small compared to the contribution of 2-pixel correlations—over 90%—to the overall redundancy of an ensemble of natural images, there has not been a detailed study of the statistical dependencies between two and three first order derivative measurements in natural images.

A further motivation for studying the statistics of gradient direction dependencies is based on authors [1, 2] who have already hypothesised that the human visual system (HVS) exploits statistical regularities in natural images to encode visual data more efficiently. This is known

as the efficient coding hypothesis and is explained in more detail in Section 1.2. There are several theoretical [18, 19], neurophysiological [3, 20] and psychophysical [21, 22, 23, 24] investigations which lend support to such a hypothesis. Moreover, it does not seem unreasonable that some form of statistical regularity in natural images could underlie Gestalt grouping rules which was first proposed in [25]. It has already been shown that when subjected to natural image stimuli, the response from pairs of (Gabor) wavelet filters—which resemble simple cell receptive fields in the mammalian visual cortex (the physiology of the HVS is summarised in Section 2.2)—show statistical relations that are consistent with the Gestalt grouping principles of collinearity and parallelism [26]. In this work, information theory is used to compute the amount of synergetic dependencies—if there are any—between geometrically meaningful measurements (gradient directions) in natural images. Synergetic dependencies would imply that there is an information advantage to a multi-local approach to the measurement of gradient directions.

1.1.7 Summary of Motivation

In summary, the motives for studying the statistical dependencies of first order derivative measurements at multiple locations and at various distances between measurements in natural images are due to:

- Statistical dependencies having already been found in natural images for pairs of edges, moreover, such dependencies are influenced by the relative orientations of the edges. Edges in an image are important because they are attributable to the properties of object surfaces and the light field in the environment. Although the study of gradient directions is not restricted—unlike edges—to locations in the image where the gradient magnitude is high.

- The orientation selectivity found in certain visual cortical neurons and the presence of lateral connections between neurons with the same orientation preference.

- The relationship between Gaussian derivatives and the receptive field profiles of visual cortical neurons.

- Low-level image processing is concerned with the extraction of features or local meaningful structures in an image (grey-scale); it turns out that it is possible to construct such feature detectors from certain combinations of derivative operators.

- The search for synergetic dependencies between multiple geometrically meaningful measurements, which can provide quantitative evidence for the advantage—in a information-theoretic sense—in adopting a multi-local approach to low-level image analysis.

1.2 Efficient Coding Hypothesis

A fundamental aim of vision research has been to understand how the human visual system (HVS) processes visual information. It is important to understand what is being measured by the HVS and how it processes these measurements. The term efficient in the efficient coding hypothesis [1, 2] is used to express that the HVS transmits the maximum amount of information from a given stimuli with the least number of neurons [3]. This is different from simply compressing an image maximally where it is necessary to be able to recreate the original image and be concerned with how much information has been lost. An efficient sensory system should exploit any redundancy found in the incoming signal. In the case of the HVS, light from the natural environment is the incoming signal; this signal is typically approximated by an ensemble of natural images in the literature.

The efficient coding hypothesis was advanced by Barlow who observed that neurons found relatively early in the processing stage of a sensory nervous system were more active, and therefore, less specific than those neurons found at later stages [2]. Barlow hypothesised that the HVS removed redundancies inherent in natural visual stimuli [2], in other words the HVS encodes the information into a more efficient form: sparse coding. Otherwise the HVS would be overwhelmed with information; it would require—if each visual stimuli were to have a unique representation—many more neurons than exist in the HVS.

For example, consider that the human retina contains around 120 million rods and about six million cones all of which can receive information from the environment. In [27], Kelly calculates that the maximum rate at which information can enter the HVS is, for the entire retina, one gigabit of data per second. This is a large amount of data and although all of this data cannot be made available for transmission, it still indicates that the HVS would be required to exploit redundancies in the signal in order to reduce the information burden without compromising the effectiveness of the HVS. Physiological evidence does suggest, as reported in Section 2.2, other cells (e.g. ganglion, bipolar, amacrine and horizontal) within the retina are combining the inputs from several of the photoreceptor cells thereby reducing the amount of

information which eventually passes through to the visual cortex via the optic nerve.

In fact, exploiting image redundancies is likely to be necessary for processes at 'deeper' levels of the visual cortex. Sparse coding represents complex data into a form that allows higher cortical areas to process the data more effectively, saving time and energy [28]. This latter point has been observed experimentally by Attwell and Laughlin who estimated that the average firing rate of neurons (< 1Hz) must be low [29]. In an extension to this work, Lennie established what would be the energy cost of an individual neuron spike, and thus, by using the known energy consumption of the cortex, Lennie was able to estimate that only 1% of cortical neurons could be active concurrently [30].

Further experimental investigations, by Vinje and Gallant, and Reiche et al., into the response properties of visual cortical neurons suggest that natural stimuli not only increase the selectivity of neurons but also decorrelates their responses [31, 32]. Moreover, the decorrelated responses to natural stimuli are not isolated to cortical neurons. For example, previous studies by Srinivasan et al. showed that the centre-surround organisation of a receptive field found in retinal ganglion cells provides a means of exploiting spatial correlations in natural images by a process of inhibition in the surround of the receptive fields [20]. It would, after all, be inefficient if the HVS had different neurons responding to the same stimuli; this would imply that more than one neuron is transmitting the same information. Therefore, neuron responses should be independent of each other, in the sense that one could not infer the response of any particular neuron given the responses of other neurons; this is a simple yet important property of sparse coding.

However, despite the encouraging arguments presented in support for the efficient coding hypothesis, the hypothesis gives no weight to what task an organism is performing. Clearly, if there is a specific task which requires additional encoding to perform, why would the HVS encode it if it is a task which is never performed or likely to be performed? Moreover, an efficient code does not necessarily mean the information contained in that code is represented in a useful way. Therefore, it seems appropriate that there must be an extension to the efficient encoding hypothesis to account for what task is being performed.

If visual systems such as the HVS sparsely code natural stimuli, evidence of this behaviour should be observed in physiological studies. Indeed the response behaviour of cortical neurons and retinal ganglion cells suggests that stimuli from natural images are being encoded in the

HVS with greater sparseness than any other class of image. To observe the effect of sparse coding, special attention must be paid to where the visual stimulus is applied. For example, a characteristic of retinal ganglion cells is their behaviour when only its classical receptive field (CRF[1]) is stimulated (experimentally this is achieved by presenting only a small patch of an image to a subject). When only the CRF of cortical cells are stimulated there is a significant chance of correlated firing, i.e. no sparsity (suppression of neuron firing). However, if the size of the patch observed is gradually increased from two to four times the diameter of the CRF, to include the non-classical receptive field (nCRF[2]), the mean spike rate falls monotonically, and the effective bandwidth of individual neurons is reduced [31, 33]. The reduction in effective bandwidth means that the range of stimuli a neuron may represent is also reduced. In spite of this, the information transmission rate increases by an amount which is statistically significant for a stimulus twice or three times the CRF diameter, but only marginally significant for a stimulus four times the diameter of the CRF. The information transmission rate is important because it should at least remain constant or increase if sparse coding is occurring, if there is a reduction, it implies information is being lost.

Observing how the information transmission rate increases or stays constant in terms of individual neuron behaviour is important because if the rate stays constant a plausible deduction is that all of the individual neurons have stayed constant. However, the findings reported in [33] suggest that it is certain neurons that are increasing their transmission rate significantly while others remain the same or even decrease significantly. Moreover, the significant increases are more frequent than significant decreases which leads to an overall rate increase. There is also an increase in the information per spike, which is calculated from the information per second divided by mean number of spikes per second. Further, in [33], it is reported that stimulation of the nCRF increases the efficiency with which neurons are processing information whereby efficiency is measured as a fraction of the amount of information transmitted divided by the bandwidth available (maximum possible transfer of information) and does not explicitly depend on the overall neuron spike rate.

In summary, it is found that the effect of contextual modulation of neurons increases i) information per spike, ii) efficiency of information transmission, and iii) selectivity (selectivity will be high when a neuron responds only to one stimulus, if it responds to all stimuli then it

[1]The size of the CRF, in Vinje and Gallant, is the diameter of the circle which circumscribes the minimum response field of the neuron [33].

[2]The non-classical receptive field lies outside the classical receptive field.

is said to have low selectivity). If we put this together with the assumption that the HVS is optimised to encode natural visual stimuli, and there have been theoretical studies—detailed in Chapter 3—by Attick, van Hateren and Field supporting this hypothesis, then it seems advantageous to study natural visual stimuli if we are to gain insight into the possible measurements and computations being performed in the HVS. All the previous points made in support of the efficient coding hypothesis provide the motivation for exploring the statistics of natural images in Chapter 3 and motivates the importance of the research carried out on the statistics of gradient directions in natural images.

1.3 Overview of Thesis

In Chapter 2, general background material is provided to cover the broad approach to vision research and introduces material of relevance to Chapter 3. Specifically in Chapter 2, the physical properties of light and the problem of inferring a 3-D scene from a 2-D image (inverse optics) is explained in Section 2.1. Both the human visual system (Sec. 2.2) and human visual perception (Sec. 2.3) are also discussed including the presentation of several vision theories: Gestaltism, unconscious inference and ecological optics. Finally in Chapter 2, a short review of computational attempts to extract features from images and the usefulness of derivative operators for the detection of features is presented (Sec. 2.4).

In Chapter 3, a more detailed critical analysis of related literature on natural image statistics is reviewed, both to motivate the work in this thesis and to provide context to the results reported.

In Chapter 4, the methodology for computing gradient measurements from images and how to compute the dependencies of gradient measurements using information theory is provided. Further, the accuracy and precision with which the dependencies are computed are also described. In Chapter 5, details on the natural image database used is provided as well as the methodology for generating the other three main image classes (phase randomized natural, whitened natural and Gaussian noise) examined in this work.

In Chapter 6, the main results of this thesis are presented on the 1-point statistics of gradient directions (Sec 6.1), and the dependencies between two (Sec. 6.2) and three (Sec. 6.4) gradient measurements. Additional experiments are reported in Chapter 7, which aim to explore the main results in Chapter 6 by analysing other image ensembles and performing further

analysis of the main results.

In Chapter 8, a methodology for how the main work in this thesis should be taken forward is outlined along with preliminary results (Sec. 8.1). Finally, brief summaries of each chapter, the main findings, and conclusions for this thesis are presented (Sec. 8.2).

Chapter 2

Background

In this chapter, relevant background material is presented in a broader context. This supplements the critical analysis of related work in the literature on natural image statistics in Chapter 3.

The nature of light is discussed in Section 2.1 and how the light field can provide information to a visual system about its environment. In Section 2.2, the physiology of the human visual system is presented, which provides support to the efficient coding hypothesis and the search for gradient direction dependencies.

In Section 2.3 the role of perception and vision theories are discussed, which motivates the search for dependencies between measurements made at separate locations in the image.

Lastly, in Section 2.4, a discussion of computational approaches to vision is given, which details the use of derivative measurements in computer vision, and how the gradient direction provides important information about the local image structure.

2.1 Light

Any visual sensory system, by definition, responds to stimulation by light. It is therefore instructive to understand the basic properties of light which enters the human visual system (HVS). After all, it is only because there can exist a consistent relationship between light and objects in the natural environment that a useful representation of the physical world can be achieved by any visual system. Objects in the natural world interact with light by absorption, reflection or transmission, therefore a brief introduction to such phenomenon are presented next.

2.1.1 Properties of Light

The largest light source in the natural environment is the sun. The sun radiates across most of the electromagnetic spectrum but significantly the intensity peak in radiation is around the middle of the visible spectrum (555nm). Light from the sun does not necessarily take a direct path to ground level. The earth's atmosphere will filter the sun's radiation, particularly ultraviolet (variations do exist across latitude) and will cause scattering, diffraction and refraction of sunlight.

Perhaps the most important or at least the largest example of scattering is that of sunlight by the earth's atmosphere. This is a result of reflections from small spherical particles and droplets which are much smaller than the wavelength λ of the scattered light[1] (Rayleigh scattering).

Light can also be scattered from particles that are larger than the wavelength of the scattered light resulting in light appearing from all directions (diffuse reflections). Thus, in the natural environment light appears not only from the sun but from the sky (skylight). An interesting difference between objects illuminated by light from the sun and the sky is its effect on the shadows of objects. Shadows originating from sunlight tend to be sharper than those caused by light from the sky. In fact, under cloudy conditions, with light coming from the sky in all directions, objects do not appear to have any shadows.

Visual systems rely on objects sending 'outward signals' which reveal properties about that object in the environment. This 'outward signal' arises from the interaction of light with objects. An incident photon from the light source strikes the surface of an object changing the trajectory of the photon. If the objects in the world were all completely transparent, i.e. all the incident light falling upon an object is transmitted through it, then the HVS would be in trouble; the reality is that all objects interact with a certain amount of light.

The surface properties of objects can cause two types of reflections: i) diffuse reflections from granular/matte surfaces where light is scattered in all directions, or ii) specular reflections from smooth polished surfaces where the angle of light reflected equals the angle of incidence (mirror-like). Both diffuse and specular reflections are idealised properties; in the real world, objects display a mixture of the two and other types as well. Moreover, the amount of diffuse or specular reflection can depend on the angle of incident light for some objects.

[1] The amount of scattering of light is inversely proportional to the fourth power of the wavelength λ^{-4} of the light. This is a reason why the sky appears blue; the reason it appears red during sunrise or sunset may be due to most of the blue wavelengths being removed as sunlight travels through more of the atmosphere.

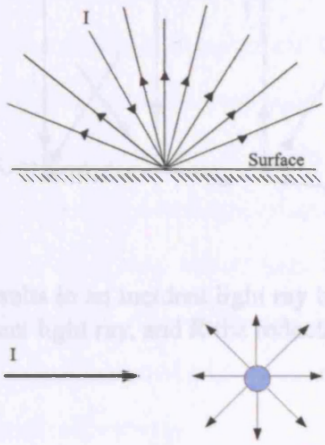


Figure 2.1: Illustrates the scattering of light rays from a surface (top) and from an object (bottom), where I indicates incident light ray. When light undergoes scattering, light rays are emitted in all directions.

In the natural environment there is not—strictly—only one source of light; reflections from objects are reflected again by other objects until finally absorbed. Hence there is generally more than one direction of light striking an object. This means that even for an idealised specular surface there can be multiple reflections and at different angles whenever other objects are present.

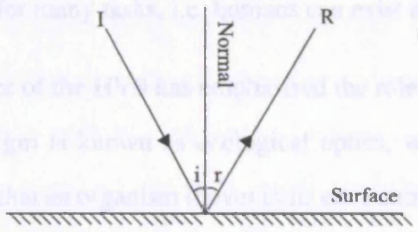


Figure 2.2: An example of a mirror reflection. I (R) indicates the incident (reflected) light ray, and i (r) is the angle I (R) makes with the Normal, where i≈r.

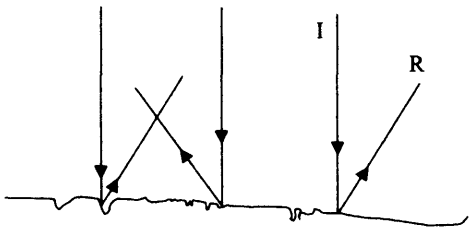


Figure 2.3: Diffuse scattering results in an incident light ray being reflected in all directions. I indicates an example of an incident light ray, and R the reflected light ray.

2.1.2 Inverse Optics

The behaviour of light when projected onto a flat surface such as the retina is important to understand because the role the eye performs in the HVS is to make perspective projections (light converges toward a single focal point) of the 3-D environment onto the human retina to form 2-D optical images. The mapping from higher (3-D environment) to lower dimensions (2-D image plane) is suited to mathematical analysis (projective geometry) as lawful relationships exist. However there are numerous ways objects interact with light (absorption, reflection, transmission) and therefore it is not possible to determine uniquely the 3-D scene from a 2-D image.

This is known as the inverse problem. How then does the HVS provide the perception of a 3-D world from a 2-D image if, for a given 2-D image, there is more than one possible 3-D scene which could be responsible. Despite this fact, the HVS appears to be adept at inferring 3-D scenes sufficiently well for many tasks, i.e. humans can exist and move in this 3-D world.

The reliable performance of the HVS has emphasised the role of optical information from the environment; this paradigm is known as ecological optics, which was first proposed by Gibson [34, 35]. He realised that an organism moves in its environment and therefore he posited that the additional information from a series of 2-D optical images could be used to further constrain a model of the 3-D world, but this still does not fully constrain the mapping from images to the environment. Effectively Gibson added the dimension of time to the problem, but time can be an added dimension to both 2-D images and the 3-D world.

Surface Orientation and Shading

A useful piece of scene information is surface orientation. This cue suggests that surface perception is a combination of the distance from an observer and the orientation of small patches of locally-flat surfaces. The two properties which are used in the literature to describe the orientation of a surface are known as slant and tilt, however, these are just a particular parameterisation where the viewer is the centre of a spherical coordinate system. More generally, the local properties of a surface are described by a depth map, which describes the depth of points relative to an observer. If the surface depth map is described as the zeroth order spatial derivative structure then the first and second order spatial derivatives of the surface depth map represent surface depth gradient and surface curvature respectively.

Both slant and tilt are concepts similar to that proposed by Marr in his 2.5-D sketch [36, 37, 38], and Marr and Nishihara [39]. For example, the optical projection of a circular disk will be an ellipse; in this coordinate system the slant determines the aspect ratio of the ellipse, and the tilt determines the orientation of the ellipse, as illustrated in Figure 2.4. Geometrically, slant is the difference in angle between the normal of a surface and an observer's line of sight and is measured perpendicular to the image plane. Tilt is the direction of the slant and is measured in the image plane. This means that on the surface of a sphere, with the eye at the centre, lines of latitude have constant slant, and lines of longitude have constant tilt.

An object illuminated by a single light source will reflect different amounts of light depending on the angle made between the light source and the surface orientation. If the normal of the local surface points toward the light source we get a maxima in the luminance pattern, but as the angle increases between the normal and the light source then less light is reflected by the surface. Thus, over the surface of the object a pattern of luminance values is created. The structure of this pattern reveals information about the local slant and tilt of the surface, if over the surface of the object the material's reflectivity does not change, i.e. we are dealing with a smooth Lambertian surface (the number of photons of light emitted per unit area is the same whatever angle one observes the object), and there is only one collimated uniform light source. In this case the total radiant power from the surface is proportional to the cosine of the angle between the surface normal and the observer's line of sight (Lambert's cosine rule).

In psychophysical experiments it has been reported that different human observers describe the local surface orientation and shape from shading qualitatively similarly when presented with

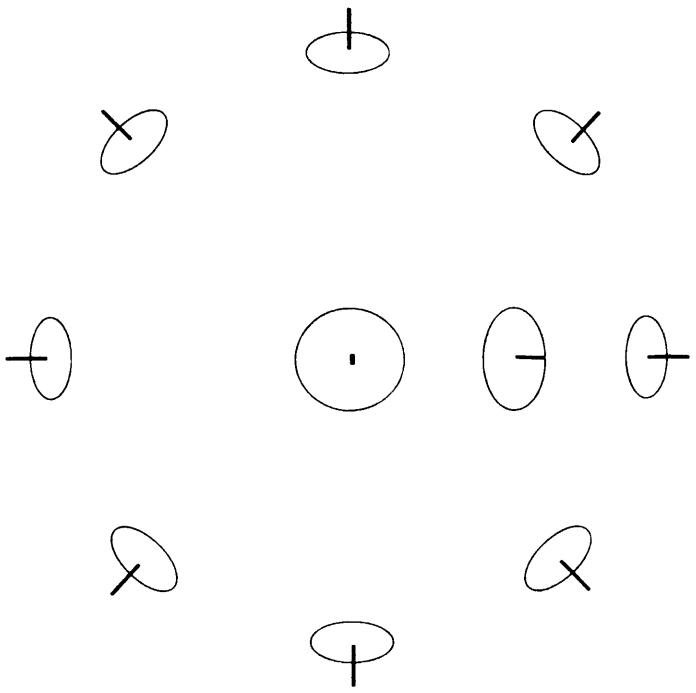


Figure 2.4: Illustration of slant and tilt. The outer ring of ellipses show constant slant but changing tilt whereas the horizontal row of ellipses from the centre to the right shows constant tilt but changing slant.

a picture of a sculpted male torso (and even when abstract sculptures are presented) [40, 41]. Moreover, without exception, each observers' results are consistent with an integral surface, i.e. a real 3-D object could be formed from their observations. However, significant variations existed amongst observers when describing the amount of depth.

If the assumption that the surface is smooth is replaced with a rough one—as is more likely to be found in the natural environment—the problem changes. A texture forms on the surface of the object, which is dependent on the illumination and provides further constraints on the object shape, which human observers can detect [42, 43]. Furthermore, it has been shown that human subjects are able to match—artificially—the light field on a rendered Lambertian sphere to the real-world illumination conditions of roughly spherical objects [44].

Depth Cues

Cues to perceiving depth in 2-D images are abundant enough so that it is not necessary to use either two eyes (binocular vision) or move in the environment, therefore depth can be perceived from 2-D images. It has been shown in psychophysical experiments that—if available—human subjects will unconsciously use previous knowledge of the sizes of objects to help determine the depth of objects [45, 46]. Whether or not previous knowledge is available, there are other

sensible assumptions which can be made by the human visual system about the behaviour of light from more distant objects in order to perceive depth.

For example, a depth cue which has a lawful dependence with distance from the observer is the exponential decay of contrast with distance. This cue occurs in scenes which contain a large range of depth for example as shown in Figure 2.5. The contrast of near objects is less than those of far objects owing to the atmosphere scattering more light from more distant objects [47]. This cue is closely linked to the depth cue caused by blurring because blurring also results in a reduction of contrast, however, it has been shown that these two effects are separate and blurring is by itself used as a depth cue by the HVS [48, 49]. Further depth cues are presented in Section A.1.

Figure 2.5: Illustration of an image where the human visual system uses the exponential decay of contrast with distance as depth cue.

2.2 Human Visual System

The physical properties and transformations of light from the environment before reaching the HVS have now been discussed in Section 2.1. A misconception is to think that the physical properties of light correspond directly to measurements made by the HVS. For example, light waves can be characterised by two physical properties: amplitude (intensity) and wavelength. However, luminance, which is measured in candela per square metre, is a perceptual quantity that incorporates the luminance function of the HVS. The luminance function is wavelength dependent, unlike intensity which is measured in watts per square metre and is independent of the observer. Moreover, the HVS is sensitive to contrast. The HVS does not perceive wavelengths

of light but rather colours. For example, there is no spectral wavelength associated with black, white, pink or brown, but these colours are perceived by the HVS. Therefore, what happens to light once it enters the human eye is discussed next.

2.2.1 Retina

When light reaches the retina the projection of the world is in fact upside down. The retina is responsible for sensing light and performing the first stages of image processing. This processing is done via a circuit of neurons which ultimately leads to signals being sent to the optic nerve (Section B.1 describes the process of neuron signals in more detail). The photoreceptor cells absorb photons which causes a chemical reaction to change the electrochemical potential of the cell [50]. However, the probability of a photon being absorbed depends on the photoreceptor's absorption spectrum and thus the energy and wavelength of the incoming photon (Further details on photoreceptor cells can be found in Section B.2).

Bipolar, Horizontal, and Amacrine Cells

Connections between photoreceptor cells and other layers of cells in the retina vary in density. The ratio of cone cells connecting to bipolar cells is $1 : 1$. This ratio is maintained for bipolar cells which connect to ganglion cells. However, further away from the fovea, the ratio changes. Progressively, more receptors connect to bipolar cells and more bipolar cells connect to one ganglion cell; so that the ratio of bipolar to ganglion cells can be $125 : 1$ [51].

Bipolar cells are either ON or OFF types, which respond differently to messages from the photoreceptor cells. An OFF pathway or ON pathway from bipolar cells through to ganglion cells is triggered depending on whether a dark image on a light background (OFF) is being observed or vice versa (ON). The ON and OFF (direct) pathways coexist with (indirect) pathways running through horizontal and amacrine cells. Horizontal cells can influence the signals sent from photoreceptor cells to bipolar cells via a feedback circuit to photoreceptor cells or directly to bipolar cells. The feedback circuitry provides a mechanism to increase the sensitivity of the photoreceptor cells in dim light conditions or reduce the sensitivity in bright conditions.

Furthermore, horizontal cells can increase the spatial sensitivity of bipolar cells. This is achieved by horizontal cells providing an opposing surround signal (centre-surround organisation). This is a process of lateral inhibition: neighbouring stimulated neurons are inhibited so that only the most strongly stimulated neuron fires thereby localizing the stimuli. However,

the cost is a reduction in sensitivity to absolute intensities. Only differences in intensities are perceived (contrast).

Amacrine cells provide interconnections between ON and OFF bipolar cells and ganglion cells which further enhances the signal sent to ganglion cells. Amacrine cells also provide the only link from rods to ON bipolar cells to ganglion cells; there is no direct link. However, this indirect link allows for the perception of very dim light via rods because the amacrine cells collect signals from many ON bipolar cells.

Overall, information through the retina follows two distinct paths: i) a direct path from the photoreceptor cells to bipolar cells to ganglion cells, and ii) an indirect path via horizontal cells and amacrine cells. However, information eventually passes through to the ganglion cells, which are then responsible for delivering messages to the optic nerve. From an information-theoretic perspective, it is worth recalling that the number of receptor cells outnumber ganglion cells by 125 : 1 [51], therefore, a substantial reduction or compression of data is occurring, which lends support to the efficient coding hypothesis. The reduction in cells is achieved via the numerous horizontal connections among the different cells in the retina layers.

Ganglion Cells

Ganglion cells deliver optical information from the retina to the brain via the optic nerve. Like bipolar cells, there are two types of cells: ON-centre and OFF-centre ganglion cells. The two types of ganglion cells were first noted in 1953 by Stephen Kuffler [52], examples of which are shown in Figure 2.8. The terms ON-centre and OFF-centre refer to how the cell responds to light. This response is characterised by the cell's receptive field whereby a receptive field is an area of the cell which, if stimulated, causes a response.

In the case of an ON-centre ganglion cell it has been found that by stimulating the receptive field near or at its centre (inner-white circle in Fig.2.8), the response of that cell is to increase its rate of firing compared with its rate when no stimulation is applied (spontaneous background firing). However, when the receptive field is stimulated away from the centre—in its surround (outer-grey doughnut region in Fig.2.8)—then activity reduces. Conversely, the OFF-centre cells have the opposite response to the ON-centre cells. These types of receptive fields are known as centre-surround and help to sharpen the spatial-localisation of boundaries in images.

Both ganglion cell types have a receptive field profile that can be modelled mathematically as a Laplacian of a Gaussian (LoG) kernel $\frac{\partial^2 G(x,y;\sigma)}{\partial x^2} + \frac{\partial^2 G(x,y;\sigma)}{\partial y^2}$. A LoG kernel looks like

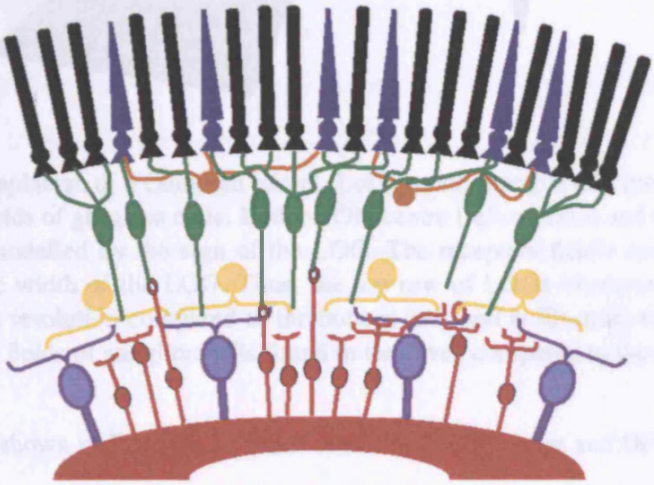


Figure 2.6: Illustrates cell layers in the retina. Light enters through the bottom and meets, in the first layer: ganglion cells—shown here with two distinct sizes (dark red and light purple). Above the row of ganglion cells are the amacrine cells, shown in yellow, which have distinctive connections running parallel to the cell layers (vertical in figure). Above the amacrine cells are the bipolar cells shown in dark green, and above them are the horizontal cells, shown in orange, running parallel (like the amacrine cells) to the cell layers. Finally the rods (black rectangular) and cones (blue triangles) are at the back of the retina (top of figure). The figure also illustrates the different 'direct' and 'indirect' paths signals take (from the top to bottom of figure) through the retina's cell layers.

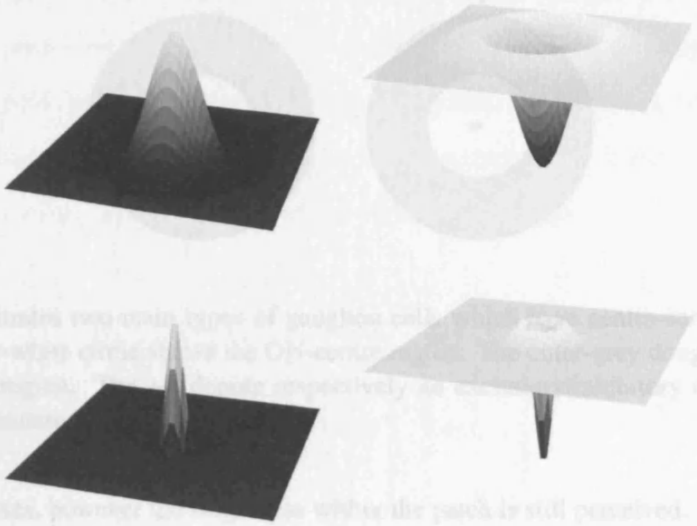


Figure 2.7: The Laplacian of a Gaussian kernel (LoG) is one example of a mathematical model of the receptive fields of ganglion cells. Both on ON-centre (left column) and OFF-centre (right column) can be modelled by the sign of the LOG. The receptive field's spatial resolution is determined by the width of the LOG. Thus, the top row of LOGs illustrates receptive fields with lower spatial resolution compared to the bottom row, and is illustrative of the difference between receptive fields of ganglion cells found in the fovea compared to the periphery.

a Mexican hat as shown in Figure 2.7, which shows both ON-centre and OFF-centre profiles. The width of the kernels are narrowed when modelling profiles of ganglion cells in the fovea, which reflects the higher spatial-resolution found in the centre of vision. However, despite the mathematical convenience and elegance of the LoG filters, retinal physiology suggests that inhibitory surrounds are weaker than modelled by the LoG filters. To complicate matters, the receptive field profiles of ganglion cells overlap, therefore, shining even a relatively small spot of light upon the retina will stimulate several ganglion cells. However, it is likely that each cell will respond differently because the same spot will fall on the centre of one cell's receptive field but upon the surround of another. Note here, that in the main results of this work (Section 6.2 and 6.4, significant statistical dependencies are found between first order derivative measurements for an ensemble of natural images).

Furthermore, another finding by Kuffler is that the centre-surround form of ganglion cell receptive fields do not respond strongly to diffuse light but do respond strongly to local spatial variation in the intensity of light falling on the cells [52]. Therefore, for a given patch of uniformly bright light the ganglion cells which respond most strongly will be those with a receptive field located on the border of the patch, i.e. at the edge. The patch interior does not

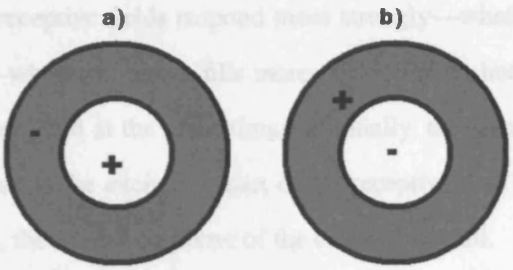


Figure 2.8: Illustrates two main types of ganglion cells which have centre-surround receptive fields. The inner-white circle shows the ON-centre region. The outer-grey doughnut represents the OFF-centre region. The +/- denote respectively an excitatory/inhibitory response to that region being stimulated.

stimulate responses, however the brightness within the patch is still perceived.

At this point in the description of the HVS, it would be necessary to move out from the retina to discuss the optic chiasm and the lateral geniculate nucleus (LGN), which link the retina and primary visual system, but these areas are of lesser relevance to this thesis. Nevertheless, for completeness, they are included in Section B.3. A summary of the primary visual cortex is discussed next.

2.2.2 Primary Visual Cortex

The primary visual cortex (V1) is a significant part of the cerebral cortex, which is a 2mm thick crumpled sheet containing 10^{11} neurons. Area V1, is the largest and most dominant visual area of the entire cortex containing about 40% of all visual neurons [30]. Moreover, 'V1 probably contains all the machinery required to account for psychophysical performance on a range of fundamental perceptual tasks—the detection of contrast patterns in brightness or color, the discrimination of orientation and spatial frequency (Vogels 1990; Zohary 1992)' [30]. Thus, V1 probably plays a role in some high-level visual tasks such as forming boundaries of objects or contour integration, however, it is unlikely that V1 is responsible for object recognition and the perception of scene layout.

V1 Cells are Orientation Specific

The receptive fields of cortical cells found in V1 are not circularly symmetric, but instead, respond to lines or edges with particular orientations. The receptive field maps, rather than simply having an ON-field and OFF-field, can have two ON-fields or OFF-fields together with one OFF-field or ON-field, as shown in Figure 2.9, the most common of which, is illustrated

in Figure 2.9a). These receptive fields respond more strongly—whether that be an inhibitory or excitatory response—when a stimulus fills more of the region, but the response is reduced if two opposing fields are filled at the same time, essentially, the responses follow the laws of addition. Therefore, if all of the excitatory part of the receptive field is stimulated along with all of the inhibitory part, the overall response of the cell will be null.

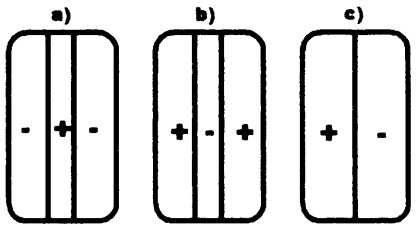


Figure 2.9: Illustrates three common receptive fields for simple cells. (+)/(-) denote excitatory/inhibitory regions. In a) a bar stimulus appropriately orientated to fit into the centre region (+) and not overlapping into any of the (-) region will cause a maximal response from the cell. Conversely, for b), a minimal response will occur when the centre region (-) is completely stimulated. In c), a maximal response will occur when a stimulus is appropriately aligned to cover all of the (+) region, and none of the (-) region.

Cortical simple cells are responsive to oriented line segments. Therefore, for a simple shape, illustrated in Figure 2.10, only the cortical cells stimulated by the borders of the shape in Figure 2.10 will be stimulated. The orientation of the contours from the border must be appropriately oriented and appropriately long to cause a response; the contour must not excite any inhibitory regions of the cell's receptive field. For example, simple cells that receive stimuli from the interior or exterior of the shape in Figure 2.10 will not be excited because no lines are present; it is an area which is smooth and uniform. However, if the shape undergoes a translation or rotation, a whole new set of simple cells will be excited. Therefore, for a given stimulus, only a small fraction of cells are excited. The properties of simple cells suggest the HVS is tuned to detecting contours and the orientation of contours. Furthermore, cortical simple cell responses are not affected by absolute intensity changes [51].

V1 Layers and Cell Grouping

In V1, neighbouring cells have receptive fields that are similar but not completely overlapping. Essentially, cells with similar properties are grouped together in the cortex; exactly what those

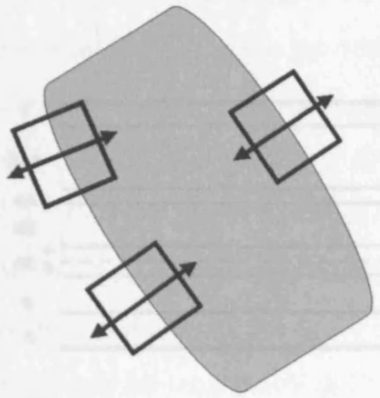


Figure 2.10: Illustrates that only cortical simple cells with receptive fields (RFs) that coincide with the boundaries of the shape will be excited. None of the cells with RFs in the interior of the shape will be excited, as the interior is uniformly bright; similarly, the exterior of the shape is uniformly bright. It is only precisely at the boundary that cortical simple cells will respond.

similarities are, and how they are spatially grouped, remains to be fully mapped out.

Broadly, however, cells receiving direct input from LGN neurons are the simplest found in the cortex. As one moves through the layers, the receptive fields of the cells go from being all centre-surround in layer $4Cb$, to some orientation-specific cells in $4Ca$, and complex cells in layers surrounding $4C$. End stop cells only occur in the deeper layers of 2 and 3, where they account for around 20% of the total cells in those layers. The changes in the OP of cells across cortical layers are not everywhere equally smooth; the OP of cells, instead of rotating 'uniformly' in a clockwise direction, may reverse direction and begin rotating anti-clockwise, as can be inferred from Figure 2.12(a). Orientation preference in terms of left or right eye dominance is discussed in Section B.4.

Lateral Connections in V1

The vast majority of connections which were initially discovered in the cortex were found to run vertically [51]—perpendicular to the orientation of the layers. However, lateral connections which can be relatively long (6mm – 8mm) [53] have been found to be more numerous. The connections run parallel to the cortex layers and connect cells with similar orientation prefer-

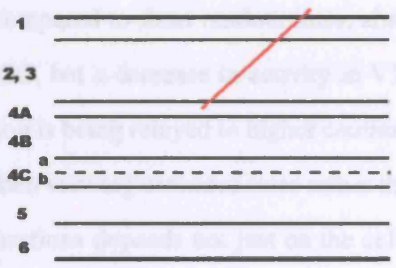


Figure 2.11: Illustrates the different layers of cells in the striate cortex. Layers 2 and 3 generally contain complex cells; layers $4A$ and $4B$ contain simple cells; layers $4Ca$ and $4Cb$ have both centre-surround and simple cells. Layers 5 and 6 contain complex cells. The red line indicates a slice through the layers from which a sample of cortex cells was taken in order to measure their OPs for Figure 2.12(a) [51].

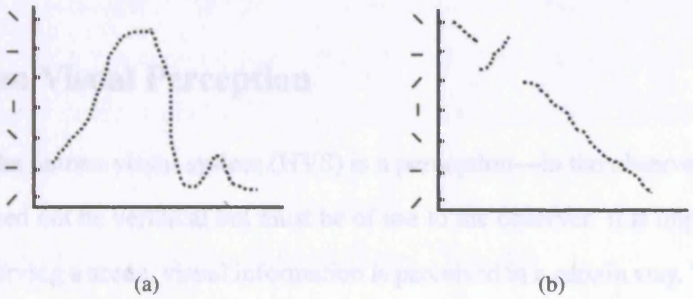


Figure 2.12: Illustrations of how the OP of cells change with distance in layers of V1. In both graphs the horizontal axis represents distance in the layers of V1. Vertical axis represents the orientation preference (OP) of the cell illustrated by a thick black line. In (a) illustrates orientation preference (OP) of cells taken from a slice through the cortex layers, as indicated by the red line in Figure 2.11, (b) is an example of another slice (not shown). Straight lines might be expected indicating that the OP of cells changes continuously with distance, but abrupt changes can be found such as a reversal in the direction of rotation with distance in (a), and 'larger' jumps in the rotation for a given distance in (b) [51].

ences. The lateral connections provide simple cells with the pathways to integrate information over a larger area than would be possible by using only the extent of their receptive fields [53].

Studies using fMRI by Kourtzi and Kanwisher [54] found that in the lateral occipital complex (LOC), located in V4, an increase in activity occurs when subjects (monkey & human) are asked to view line objects compared to short random lines; similarly, Murray et al. [55] found an increase in activity in LOC, but a decrease in activity in V1. The fact there is a change in V1 suggests some information is being relayed to higher centres of the visual system by cells in the primary visual cortex when viewing extended lines rather than random short line segments. The strength of lateral connections depends not just on the cells having parallel OPs, but also the physical spatial relation between the receptive fields [56, 57]. Thus, the strongest lateral connections occur when the cells have parallel OPs as well as being co-linear spatially, and on the contrary, a less strong connection will occur when parallel OPs are perpendicular to each other [58, 59].

The lateral connections between cells is important, especially from the viewpoint of perception and the global integration of local visual cues. Exactly how such connections fit in with the topographic nature of the visual cortex is an unsolved problem. Vision scientists are thinking of ways to exploit the concept of lateral connections, but the fact any long-range global integration is taking place as early as V1 tells us how important integration of visual information will be to understanding human visual perception, which is discussed next.

2.3 Human Visual Perception

The output of the human visual system (HVS) is a perception—in the observer's mind—of the world which need not be veridical but must be of use to the observer. It is important to discuss why, when observing a scene, visual information is perceived in a certain way. There are several vision theories which try to address this problem: ecological optics, which emphasises the role of the environment and learning to perceive the world through our movement (interaction) within it. Another approach simply posits that the world appears a certain way as a result of the way the HVS is wired: if the system were wired differently the world would appear differently. It is still a controversial issue as to how much of perception is through bottom-up processing in the HVS: is the retinal image (the input data of the world) modified into higher levels of representations? How much is top-down: is the input the higher level representation which is

modified by the image data?

Even though early low-level processes and behavioural responses are likely to be bottom-up (feed-forward), as demonstrated by rapid categorisation experiments [60]. In [60], subjects were asked to respond to whether an animal was present or not in an unseen image that was flashed for 20ms. The response times were such that only a feed-forward mechanism could have been utilised by the HVS given the limitations on the transfer rates of information in neurons. Nonetheless, there are examples of where top-down mechanisms are contributing to certain visual percepts, as demonstrated later in this section.

What is certainly a relevant question for the vision scientist is why, in the words of Wertheimer, "When we are presented with a number of stimuli we do not as a rule experience 'a number' of individual things, this one and that. Instead, larger wholes separated from and related to one another are given in experience" [61]. In other words, when observing a scene, whole objects are perceived, not an array of separate picture elements (pixels).

Wertheimer was one of the co-founders of Gestalt theory (the other founders were Koffka and Köhler). A theory which has had an enormous impact on research in visual perception in both the twentieth and beginning of the twenty-first century. The Gestaltists made a significant contribution to understanding perceptual organisation by establishing a number of qualitative grouping rules which govern human visual perceptions. Perceptual grouping can be described as a process by which our visual system collects data from an image and organises that data into subsets. This paradigm is the opposite of structuralism, which was proposed by Wundt, and states that visual perceptions are built through associations between visual atoms, i.e. local analysis of a scene contains sufficient information to build perceptions. Gestalt theory opposed this view and proposed that global integration of local analysis must also be performed.

2.3.1 Unconscious Inference and Likelihood Principle

Unconscious inference incorporates concepts from both structuralism and Gestaltism, but it is in contradistinction to Gibson's ecological optics theory. It was proposed by Helmholtz in order to explain how the HVS is able to interpret 2-D retinal images to form a 3-D world. The explanation given is known as the likelihood principle; it states that the HVS bases its interpretations on the most likely 3-D environment which could have caused the 2-D retinal image [62].

The theory of unconscious inference can be cast in a Bayesian framework: the HVS

may adopt a Bayesian approach to make inferences from images (evidence) about the 3-D world [63]. This can be formulated in the following way. Given an Image I, there exist a number (n) of possible scene descriptions for the 3-D environment E_n. First, assign prior probabilities to an image I by $P(I)$ and scenes E_n by $P(E_n)$. Second, assign conditional probabilities on scenes E_n given I by

$$P(E_n|I) = P(E_n, I)P(I), \tag{2.1}$$

and an image I given scenes E_n by the likelihood function:

$$P(I|E_n) = P(I, E_n)P(E_n). \tag{2.2}$$

Thus, Bayes theorem is formulated from equating the joint distributions $P(E_n, I) \equiv P(I, E_n)$ to give, for a particular scene $E_{n=1}$:

$$P(E_1|I) = \frac{P(I|E_1)P(E_1)}{P(I)} \tag{2.3}$$

where $P(I) = \sum_n P(E_n)P(I|E_n)$ is a normalisation factor, i.e. what is the probability of I given all the possible scenes E_n. Equation 2.3 is of the form:

$$\text{(posterior)} \propto \text{(likelihood)} \times \text{(prior)} \tag{2.4}$$

The likelihood principle implies that the HVS has a prior estimate of its belief in a certain hypothesis $P(E_n)$ without evidence, but after evidence has been observed, the HVS recalculates a posterior belief in the hypothesis to form $P(E_n|I)$.

The likelihood principle provides a framework in which it is possible to use image statistics to make a guess at the prior distributions used in Equation 2.3. For instance, $P(I)$ could be a prior distribution of a statistic found from analysing natural images. $P(I)$ can also provide information about $P(E_n)$, which serves as a prior model of the environment. Furthermore, the likelihood principle provides a framework in which competing hypothesis and ambiguous scenes can be handled; this phenomenon is called multistability, which is described in Section 2.3.4 and by Figure C.5.

Inferring a scene from images remains an important concept in computer vision (as described in Sec. 2.4) for the development of computational algorithms and it is hypothesised that

the HVS uses inference to form percepts. This means that the HVS is prone to errors. If the assumptions used by the HVS are false this can lead to wrong scene interpretations from a set of retinal images.

2.3.2 Gestaltists

Gestaltists studied how certain images evoke erroneous percepts and examples of these illusions will be shown. The view of Mach: 'The visual sense acts therefore in conformity with the principle of economy, and, at the same time, in conformity with the principle of probability, when it exhibits a preference for straight lines.' [62] incorporates the principle of Prägnanz proposed later by Gestaltists and suggests the need for inference in visual perception. Prägnanz is a principle where the interpretation of a visual scene takes on the simplest possible explanation; this is related to the 'goodness of figures'.

Figural goodness underpinned much of Gestalt thinking. The claim is that the HVS prefers regularity and simplicity when it comes to the perception of figures. For example, a circle would be preferred to a random distribution of lines. This means that the Gestaltists thought several perceptions were possible for any given figure, but the figure which is most 'good' is the one perceived. Figural goodness tried to address the phenomenon of competing hypotheses, in an analogous way to the Bayesian approach which is to perceive the most likely figure. In other words, the Gestalt theory of figural goodness is that the HVS has a preconceived notion of what is 'good' and this could be interpreted within a Bayesian framework as some kind of prior information or model. Unfortunately, the Gestaltists only gave qualitative descriptions of what 'goodness' might be. This became a criticism of work by the Gestaltists, what is 'good'? Such problems and other flaws in Gestalt theory are discussed in Section C.1.

In spite of these reservations, the Gestaltists made significant contributions to vision through psychophysical experiments and determined qualitative grouping rules that govern how visual data is partitioned. The main principles identified were similarity, proximity, good continuation, and closure [61, 64]. Later, additional grouping rules such as common region [65] and connectedness which may be linked to proximity [66], were also proposed. However, the overall theme of the Gestalt paradigm was holism: "the whole is greater than the sum of the parts". Further details on holism are provided in Section C.2.

Perceptual Grouping

Perceptual grouping has an ambiguous meaning, but herein it shall be described as the process of grouping visual information into distinct objects. Moving from smaller, separated units of structure to larger organised segments of structure. An alternative view to perceptual grouping might be the process by which our visual system collects data from an image and partitions them into particular 'arrangements'. However, the partitions are unknown, the format of the data is unknown, but if the 'arrangements' can be understood, this may lead to an understanding of how to perform perceptual grouping. The Gestaltists addressed this problem and proposed several 'principles' of grouping which are explained in Section 2.3.3. The necessity for incorporating 'principles' of grouping or some priors of how to organise image data is illustrated by Figure 2.13 which shows the enormous computational task of partitioning data into subsets by demonstrating how many ways the integer n can be written as a sum of positive integers.

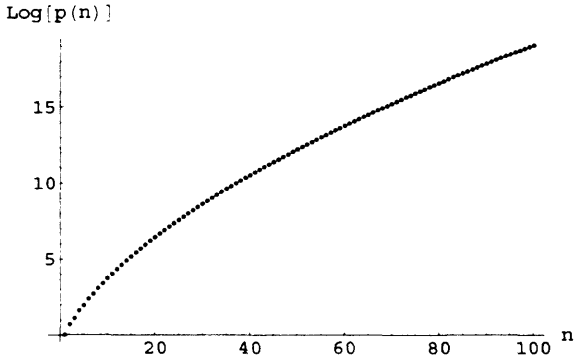


Figure 2.13: Shows the partition function, $p(n)$, on a Logarithmic-normal scale. The function, p(n), demonstrates the number of ways the integer n can be written as the sum positive of integers, e.g. $n = 4$, can be written as any one of the following subsets: {3+1}, {2+2}, {1+1+1+1}, {2+1+1}, {4}. The rise with n in the number of subsets increases asymptotically like $e^{\sqrt{n}}$ [67].

2.3.3 Gestalt Principles of Grouping

The four main grouping principles identified by the Gestalts are i) similarity, ii) common fate, iii) good continuation, and iv) closure. Only similarity will be described here, and the other principles are described in Section C.4).

Within the Gestalt principle of similarity are different similarity cues that may compete with each other. For example, the grouping principle of similarity implies that objects with similar properties—such as size (Fig. 2.14a), orientation (Fig. 2.14b), and colour (Fig. 2.14c)—will be grouped together when all other principles are equal. Furthermore, proximity—a grouping

cue—is linked with similarity of position, as illustrated in Figure 2.14(d).

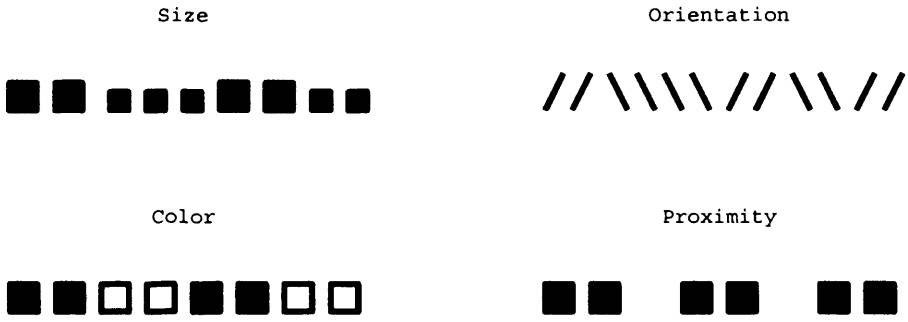


Figure 2.14: Top left: squares of the same size are grouped together. Top right: lines of the same orientation are grouped together. Bottom left: squares of the same colour are grouped together. Bottom right: pairs of squares closest to each other are grouped together.

Although proximity is regarded as the most fundamental Gestalt grouping principle [68] because of the HVS's acute sensitivity to it [69], a criticism of proximity is its failure to describe how the strength of the cue diminishes with distance between elements. In [70], Oyama found that when presenting a regular rectangular dot lattice with vertically and horizontally aligned dots, which gives rise to two different (vertical and horizontal) organisations, the length of time an observer experiences one organisation over the other was found to have a power law dependence on the ratio of distances between the vertical and horizontal elements in the organisation. Furthermore, Kubovy and Holcombe [68] were able to show a mathematical relationship between the probability of perceptually organising the lattice in the horizontal or vertical directions and the minimum distance between the dots in the lattice. Moreover, Kubovy and Holcombe found that the longer the stimulus is visible the stronger the proximity cue becomes and that scale invariance is observed: increasing the horizontal and vertical scales equally did not change the strength of the cue. However, the proximity cue described by these results follow an exponential rather than a power law (i.e. scale-invariant) dependency on distance. Moreover, later studies found that a probabilistic Gaussian distribution model fitted the proximity cue well [71].

From the description given so far for the principle of similarity, it is not known, for example, how the cue of proximity supersedes similarity of intensity? Details of the competition between grouping rules, and extensions to the grouping rules are provided in Section C.7.

2.3.4 Global Versus Local Approach to Perception

The local edge-based approach, as described later in Section 2.4, to region segmentation is in contradiction to the holistic, global philosophy of the Gestaltists. The Gestaltists believed perceived regions arose because of their common properties attracting them together, rather than first performing some local edge detection. In support of Gestalt theory, there could be no better examples than illusionary contours and amodal completion, an example of which is illustrated Figure 2.15. The fact the HVS performs amodal completion lends support to the importance of global analysis in perception over just local analysis. Even 3-D volumetric illusory percepts of spheres can be produced [72]. This is an important result because it suggests that illusory contours are not merely some minor add-on to an edge-detection mechanism, but rather an instance of a more general perceptual phenomenon.

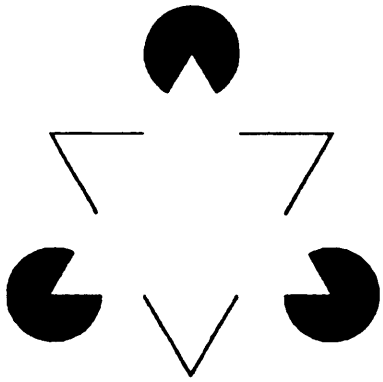


Figure 2.15: The Kanizsa triangle is a good example of how a complex set of local cues give rise to a global phenomenon despite it being, in this instance, an illusionary phenomenon.

The explanation offered by Gestaltists for the types of figures like the Kanizsa triangle (Fig. 2.15) is that objects which are partly occluded are 'filled in' by the 'simplest' figure. This refers back to the principle of Pragnänz: what is perceived will be as simple and regular as possible or 'good'. In the Kanizsa figure, it is thought that the 'pac-men' are more likely to be full circles; this gives rise to the illusion that a white triangle lies above the circles.

Another type of illusion, which Gestaltists tried to explain, is multistability. Multistability is the term used when the HVS experiences a perception that alternates between different possibilities. This phenomenon is illustrated in Figure 2.16. Depending on whether the white part is the background or foreground the percept is altered. A further discussion of multistability can be found in Section C.5.

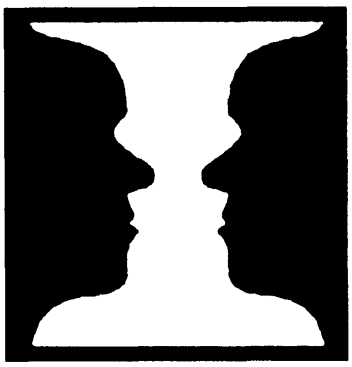


Figure 2.16: An example of multistability; the perception alternates between a white vase and two black faces. The problem is also experienced with figure-ground scenes: depending on whether you perceive the white part as the background or the foreground, your perception will change.

Conclusions

Overall, the Gestaltists made a significant contribution to theories of visual perception by realising that perceptual tasks need to incorporate more than just local analysis, and perhaps some form of multi-local analysis.

Computational algorithms have incorporated Gestalt theory in their approach to region segmentation by adopting a global analysis of a scene rather than just local analysis. For example, rather than detect (local) edges to segment an image—as in the Canny edge detector—Shi and Malik [73] used a graph theoretic approach, which utilises the Gestalt grouping principle of similarity to segment the image into meaningful regions. Essentially, this becomes a problem of graph partitioning. The algorithm groups pixels together based on maximising both the similarity and dissimilarity between sets of pixels and where the similarity measures are based on the luminance, colour and texture of the pixels (motion is also implemented where images are treated as a spatiotemporal data set and pixels that are in the local spatiotemporal neighbourhood are connected together). The results seem to outperform the Canny edge detector. Thus, from a conceptual point of view, the advantages or at least the possibility of using techniques that make use of Gestalt theory and global analysis should be evident.

Furthermore, there have been successful attempts made at probabilistic models that utilise Gestalt grouping cues to group the edge points of objects into extended chains within artificial scenes[2]. Developments have been made by incorporating the statistics of natural image contours as priors within a Bayesian framework [74]. The success enjoyed by such approaches helps to

[2]Trying to create contours which bound an object of arbitrary shape within a natural scene has proved to be a much harder task.

motivate the studies of natural image statistics which have received constant attention for over half a century. Therefore, natural image statistics are important, and the main findings in the literature are covered in Chapter 3.

2.4 Computational Approach to Vision

Computational theories can approach problems of image processing without need for recourse to the HVS, but it is still instructive for either physiologists or computational theorists to understand some of the work in either field. For example, an autonomous robot which can 'see' may not measure the environment with the same receptors or utilise the same image processing algorithms as the HVS but ultimately the robot should be able to perform visual tasks as competently as the HVS or better.

The interdisciplinary approach to vision between computer vision and human visual perception increased after Marr [38] proposed that visual perceptual tasks should be expressed as nothing more than an information processing task. In Marr's scheme, the theory of vision has three levels:

1. Computational level which describes the task,

2. Representational and algorithmic—what form does the input and output take and what transformations are performed on the input to create an output, and

3. Implementation level—the hardware through which the representations and algorithms are implemented, e.g. a computer.

2.4.1 Image Features

Much of low-level image processing is concerned with the extraction of features or local meaningful structures in grey-scale images. Four low-level feature types proposed by Marr [38] were edges, bars (short line segments which end outside the receptive field), blobs (short bars with terminations at both ends or more commonly as a brighter/darker region with a darker/brighter surround), and terminations (the end of edges or bars). These low-level feature types are all qualitatively different and make up the elements in his concept of the raw primal sketch[3].

[3]The feature types described by Marr had been previously identified in the works of Julesz as the atoms of early vision. However, Julesz placed the feature types in the context of modelling texture (texton theory) in images and texture perception [75] whereby texture is defined as a region with features that share some common statistical properties [21].

Within each type is a token which corresponds to a particular type, furthermore, each particular type is instantiated by certain properties such as its position, size, orientation and contrast. The raw primal sketch aims to have a direct interpretation—in a computational sense—of what processes might be occurring in the HVS, especially in V1. Indeed the physiological findings in the striate cortex of the cat by Hubel and Wiesel [76, 77] appear to have influenced the work of Marr. For example, the end-stop cells described by Hubel and Wiesel have a direct analogy to Marr's inclusion of bars in his set of image primitives (types). Furthermore, the edge operators proposed by the Marr-Hildreth algorithm [78], presented in Section D.1, bear resemblance to certain cortical visual neuron receptive fields, and the detection computation as the sum of operator outputs bears resemblance to theories proposed by Hubel and Wiesel on what simple V1 cells are doing to detect edges.

2.4.2 Derivative Operators as Feature Detectors

Before discussing how meaningful feature detectors can be constructed from certain combinations of image derivatives, it is necessary to introduce scale space theory because an image of the form $L(x, y)$ is a discrete function, and therefore applying a differential operator directly to $L(x, y)$ is ill-posed. However, if then the image is modelled as a manifold (set of lines, curves, and surfaces), in particular an infinitely differentiable manifold C^{∞}, a whole branch of mathematical tools (differential geometry) becomes available for analysing images. The manifold allows functions to be defined which are continuous and differentiable, and to define a local coordinate system.

In order to move from a discrete representation of an image function $L(x, y)$ to a continuous related set of derived images, a scale parameter is introduced, which is a reasonable step given that $L(x, y)$ is an observable measured through some aperture of finite resolution. Moreover, features exist at certain scales in the image, and therefore it is necessary to probe the image at different scales. To do this, the original image $L(x, y)$ is embedded in $L'(x, y; \sigma)$, where σ is the resolution parameter and $L'(x, y; \sigma = 0) = L(x, y)$ [79].

Physically, an image is the result of a real measurement made by a camera. A camera records a scene and that camera will have a label stating its resolution in terms of a certain number of pixels. The size of one of these pixels (assuming they are all the same) is what determines the inner scale of the camera (σ): the smallest details which can be resolved by the camera from the scene (it is not infinitesimal, i.e. $\sigma = 0$). The total array of pixels is then

the largest detail which can be resolved by the camera and represents its outer scale. Within each pixel, light is collected over a certain area (the size of the pixel). Thus, the amount of light falling on a pixel is integrated over space. A similar process occurs in the HVS, the eye is performing a measurement with limited resolution determined by the photoreceptor cells (cones and rods) in the retina; however, the resolution is reduced by the fact that receptive fields in the HVS collect inputs from several photoreceptor cells.

Now that the scale space paradigm has been introduced, it is possible to continue with the discussion of the relationship between derivative operators and images. The order of the derivative measurement applied to an image can provide information on its local structure. For example, the first order derivative calculates the slope, and the second order derivative the curvedness of the local intensity landscape in the image. This mathematical analysis of local structure is well-known as a Taylor expansion. In 2-Dimensions this is a essentially a series of terms which approximate how the surface changes if we move from the current location (x, y) an infinitesimal amount $(\delta x, \delta y)$. When more higher order derivative terms are included in the series the better the approximation of the local structure (ignoring the structure of noise in the image), and for an infinite number of terms, the series becomes an exact description. The series in Equation 2.5 shows the Taylor series up to second order derivative terms, for brevity all $L(x, y)$ on the right hand side of the equation are written as L:

$$\begin{aligned} L(x+\delta x, y+\delta y) = L &+ \frac{\partial L}{\partial x}(x-\delta x) + \frac{\partial L}{\partial y}(y-\delta y) \\ &+ \frac{1}{2!}\left(\frac{\partial^2 L}{\partial x^2}(x-\delta x)^2 + 2\frac{\partial^2 L}{\partial x \partial y}(x-\delta x)(y-\delta y) + \frac{\partial^2 L}{\partial y^2}(y-\delta y)^2\right) \\ &+ O(\delta x^3, \delta y^3) + \ldots \end{aligned} \tag{2.5}$$

It is possible to demand that this description of the local image structure is independent of the choice of coordinate system. In other words, the quantities measured are invariants, i.e. they do not change under transformations such as rotations and translations of the image. Invariants tend to be associated with physical entities because a physical entity is not dependent on the abstract coordinate system which is imposed. Therefore, invariants do not change under transformations of a coordinate system.

Two examples of invariants of image structure—composed from first order derivative measurements—are the gradient magnitude and the gradient direction; the gradient field of

an image is given by:

$$\nabla L(x,y) = \begin{pmatrix} \frac{\partial L(x,y)}{\partial x} \\ \frac{\partial L(x,y)}{\partial y} \end{pmatrix} \equiv \begin{pmatrix} L_x \\ L_y \end{pmatrix} \tag{2.6}$$

where $\nabla \equiv \left\{\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right\}$. The gradient magnitude is then calculated from $\sqrt{\left(\frac{\partial L}{\partial x}\right)^2 + \left(\frac{\partial L}{\partial y}\right)^2}$ and the gradient direction from $\arctan\left(\frac{\partial L_y}{\partial L_x}\right)$.

2.4.3 Gauge Coordinates and Gradient Direction

The gradient direction $\hat{w}$ points along the direction of greatest change in the image intensity. Thus, it provides information about the extrinsic geometry of the local image structure, and along with the direction perpendicular to it $\hat{v} = \hat{w}_\perp$, which is tangential to the isophote (curves connecting points of equal intensity, therefore the gradient direction is normal to isophote), provides a means to establish first order gauge coordinates. Problems arise when the gradient magnitude is zero, i.e. the gradient direction is undefined and thus so are the gauge coordinates. These locations are known as stationary points or singularities and there are different classes such as saddle points, and minima and maxima in the intensity. In Figure 2.17, the $\hat{v}$ and $\hat{w}$ gauge frame in each pixel is shown plotted for a first order mixed partial derivative of a Gaussian blob.

2.4.4 Corner Detectors

A corner is a local neighbourhood in the image which has two edges with different orientations that intersect each other. A common form for corner detectors [80, 81, 82, 83, 84, 85, 86, 87, 88] can be described by finding the local spatial extrema in $L_{vv}L_w^n$, which is commonly recast in terms of the product of isophote curvature κ multiplied by the gradient magnitude raised to some power n: $\kappa \equiv -\frac{L_{vv}^n}{L_w}$. L_{vv} is the second derivative in the direction perpendicular to the gradient direction and is invariant to monotonic transformations of the image intensities.

$$L_{vv} = \frac{L_x^2 L_{yy} + L_y^2 L_{xx} - 2L_x L_{xy} L_y}{L_x^2 + L_y^2}. \tag{2.7}$$

The particular case of a corner measure proposed by Blom [85] described by Equation 2.8 gives an affine invariant corner detector:

$$\kappa L_w^3 = L_{vv}L_w^2 = L_x^2 L_{yy} + L_y^2 L_{xx} - 2L_x L_{xy} L_y. \tag{2.8}$$

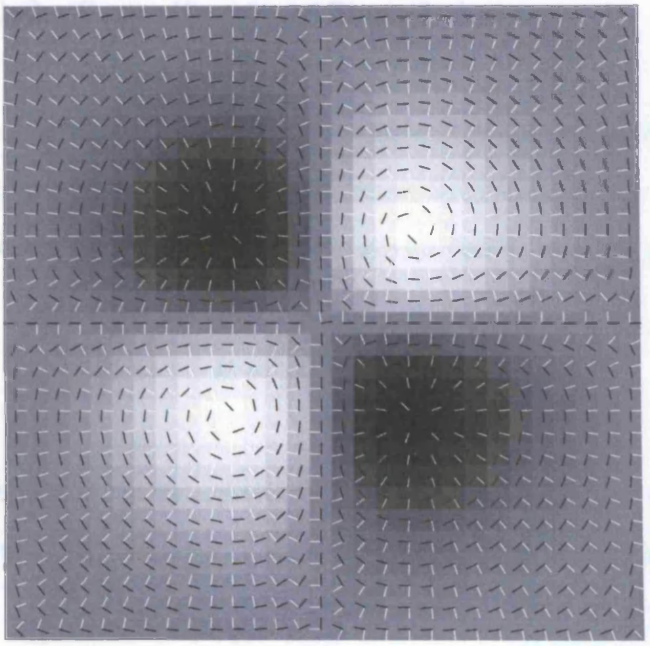


Figure 2.17: Each pixel contains an orthogonal frame composed of the gradient direction $\hat{w}$ (white arrows) and the gradient direction rotated clockwise by $\pi/2$ $\hat{v}$ (black arrows), which are respectively perpendicular and tangential to the curves of equal intensity in the image (isophotes).

Although such corner operators are susceptible to errors from noise in the image, a further optimisation is obtained by considering the signed scale space maxima (positive maxima, negative minima) [89] of the corner operator across several scales. The consideration of scale is important to detect corners with different spatial extent and helps to reduce the problems with noise from derivative measurements. A sharp corner produces a strong response if a fine scale operator is used, and conversely a broad (rounded) corner produces a strong response when a coarse scale operator is used.

A different approach to corner detectors of the form $L_{vv}L_w^n$ is one based on the gradient direction [90]. In [90], the corner measure used is $\|\nabla\theta\|_2$, the 2-norm of the grad of the gradient direction computed from $\theta(x,y) = \arctan\left(\frac{\partial L_y}{\partial L_x}\right)$:

$$\begin{aligned}
\nabla\theta &= \{\theta_x, \theta_y\} \\
&= \left\{ \frac{L_x L_{xy} - L_{xx} L_y}{L_x^2 + L_y^2}, \frac{L_x L_{yy} - L_{xy} L_y}{L_x^2 + L_y^2} \right\} \\
\|\nabla\theta\|_2 &= (\theta_x^2 + \theta_y^2) \\
&= \frac{(L_{xy} L_x - L_{xx} L_y)^2 + (L_{yy} L_x - L_{xy} L_y)^2}{(L_x^2 + L_y^2)^2}.
\end{aligned} \tag{2.9}$$

2.4.5 Second Order Gauge Frame and Shape Index

It is possible to set up orthogonal directions of curvature which depend on the local intrinsic image geometry which can be used to form a local frame. Such directions are called the principle directions. To derive the principle directions, we take the derivative of the gradient (Hessian):

$$\mathbf{H} = \begin{pmatrix} L_{xx} & L_{xy} \\ L_{yx} & L_{yy} \end{pmatrix} \tag{2.10}$$

and decompose the matrix into its eigenvectors and eigenvalues. To find the eigenvectors we first need to find the two eigenvalues κ_1, κ_2 by solving $|\mathbf{H} - \kappa\mathbf{I}| = 0$ where $\mathbf{I}$ is the identity matrix, this gives:

$$\begin{aligned} \kappa_1 &= \frac{1}{2}\left(L_{xx} + L_{yy} - \sqrt{4L_{xy}^2 + (L_{xx} - L_{yy})^2}\right) \\ \kappa_2 &= \frac{1}{2}\left(L_{xx} + L_{yy} + \sqrt{4L_{xy}^2 + (L_{xx} - L_{yy})^2}\right). \end{aligned} \tag{2.11}$$

It is then possible to find two eigenvectors of H by solving $(\mathbf{H} - \kappa_1\mathbf{I})\mathbf{K}_1 = 0$ and $(\mathbf{H} - \kappa_2\mathbf{I})\mathbf{K}_2 = 0$ to get:

$$\begin{aligned} \mathbf{K}_1 &= \begin{bmatrix} \frac{L_{yy} - L_{xx} + \sqrt{4L_{xy}^2 + (L_{xx} - L_{yy})^2}}{2L_{xy}} \\ 1 \end{bmatrix} \\ \mathbf{K}_2 &= \begin{bmatrix} \frac{L_{xx} - L_{yy} + \sqrt{4L_{xy}^2 + (L_{xx} - L_{yy})^2}}{2L_{xy}} \\ 1 \end{bmatrix}. \end{aligned} \tag{2.12}$$

An example of the local frames of K_1 and K_2 plotted for each pixel of an image is shown in Figure 2.18.

Not only do the eigenvectors of the Hessian have a significant geometric interpretation but so do its eigenvalues κ_1 and κ_2 which are known as the principle curvatures and describe the amount of curvature. In fact, the relationship between the principle curvatures provides a description of the shape of the local image surface which was introduced by Koenderink along with his definition for curvedness. Curvedness is defined as

$$\frac{1}{2}\sqrt{\kappa_1^2 + \kappa_2^2} \tag{2.13}$$

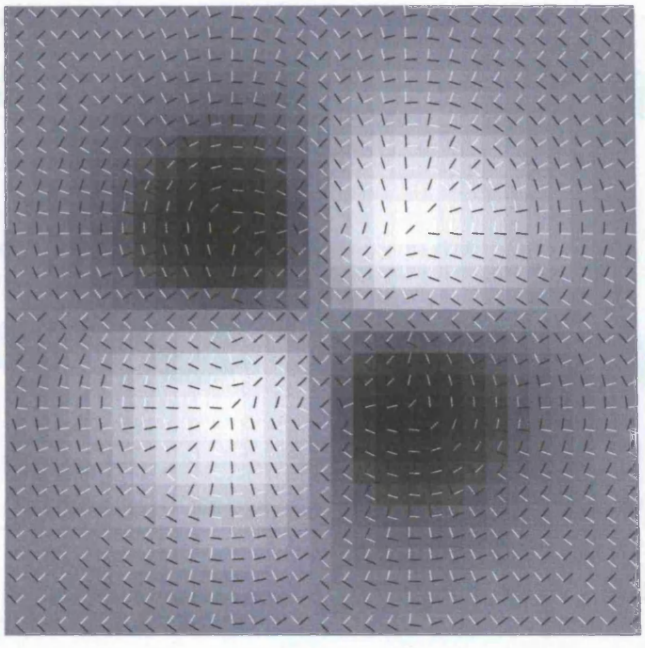


Figure 2.18: Each pixel of this image—a mixed first order derivative Gaussian blob—contains an orthogonal frame of principal directions K_1 (white) and K_2 (black) which show the maximal and minimal curvature directions respectively.

and in terms of partial derivatives as

$$\frac{1}{2}\sqrt{L_{xx}^2 + 2Lxy^2 + L_{yy}^2}. \tag{2.14}$$

The shape description, known as the shape index, is defined as

$$\frac{2}{\pi}\arctan\left(\frac{\kappa_1 + \kappa_2}{\kappa_1 - \kappa_2}\right) \tag{2.15}$$

where $\kappa_2 \geq \kappa_1$, and in terms of partial derivatives,

$$-\frac{2}{\pi}\arctan\left(\frac{L_{xx} + L_{yy}}{\sqrt{4L_{xy}^2 + (L_{xx} - L_{yy})^2}}\right). \tag{2.16}$$

The shape index has the range $[-1, +1]$. Examples of the shapes are shown in Figure 2.19, and an example of the shape index of a natural image is shown in Figure 2.20.

The shapes can be further generalised from the relative signs of κ_1 and κ_2 and whether one or both are equal to zero. For example, a concave shape occurs for $\{\kappa_1, \kappa_2 > 0\}$, saddle $\{\kappa_1 > 0, \kappa_2 < 0\}$ or $\{\kappa_1 < 0, \kappa_2 > 0\}$, convex $\{\kappa_1, \kappa_2 > 0\}$, spherical $\{\kappa_1 > 0, \kappa_2 > 0\}$ or

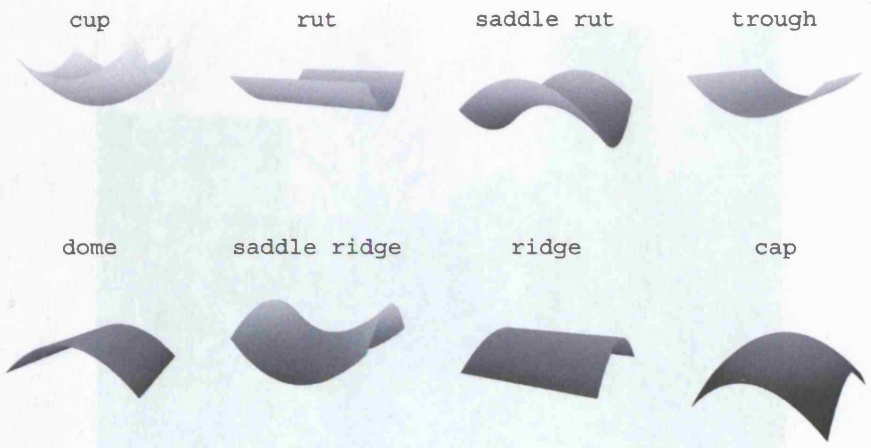
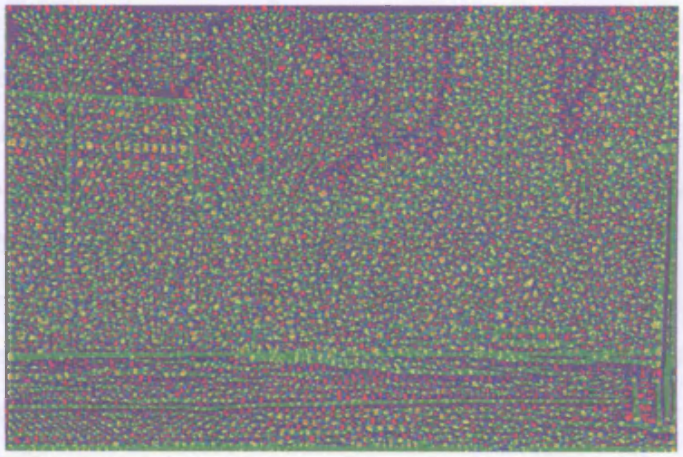


Figure 2.19: Illustrates eight shapes from the shape index defined by Koenderink.

$\{\kappa_1 < 0, \kappa_2 < 0\}$, cylindrical $\{\kappa_1 = 0, \kappa_2 \neq 0\}$ or $\{\kappa_1 \neq 0, \kappa_2 = 0\}$, and flat $\{\kappa_1 = 0, \kappa_2 = 0\}$.

(a)



(b)

Figure 2.20: Example of a natural image (top image) and its shape index (bottom image) encoded using the scale shown at bottom.

Chapter 3

Natural Image Statistics

In this chapter, a critical analysis of related literature to this thesis is presented. The assumption that the HVS is optimised to encode natural visual stimuli is based on the premise that given the set of all possible images, natural images are sparse. If it is assumed that the natural environment can be approximated by an ensemble of natural images, such an ensemble should contain statistical regularities that distinguish it from other image classes. This is not to say that within the subset of images, described as natural, there is not a degree of variability. There is variability and the HVS would not want to be insensitive to such differences: there is no trouble in distinguishing a forest from a crop field. Research on natural image statistics has, by studying—predominately—natural image ensembles rather than individual natural images, attempted to characterise the properties of any regularities.

Regularity in the mean power spectrum of an ensemble of natural images (Sec. 3.3) is presented first, including the scale-invariant properties of natural images in Section 3.3.1. In Section 3.4, the contrast statistics of natural images are described, but these regularities do not distinguish an ensemble of natural images from other image classes. In Section 3.5, higher order statistical regularities in natural images are discussed; for example, the statistics of distributions of filter responses (Sec. 3.5.1), phase spectra (Sec. 3.5.2), and dependencies between edges (Sec. 3.5.4). The psychophysical significance of any regularities found are also discussed in Sections 3.3.4 and 3.5.3.

3.1 Introduction

A fundamental aim of vision research has been to understand how the human visual system (HVS) processes visual information. Two of the key questions are i) what is being measured by the visual system? and ii) under what system are those measurements being processed? In

order to help solve such questions, it is useful to consider constraints on the evolution of the HVS. The first assumption is that the visual system must enable the species to perform essential tasks that are necessary for its survival. The second is that the system's development has been influenced by the most common signals it receives. Thirdly, the hardware of the system—how much information can be transferred and how quickly through the system.

The last two assumptions have motivated research into the statistics of natural scenes because this is the environment in which the HVS has evolved. Furthermore, understanding the amount of information contained in natural scenes will help establish what the information burden is on the HVS. At this point it may seem that the only benefit of finding a relationship between the statistics of natural images and the processes of the HVS is one of knowledge, i.e. understanding such a relationship will help establish what functions neurons are performing and how the environment has influenced the properties of neurons. However, there are two further benefits for computer vision. One is the development of computational algorithms which can perform vision tasks, and another one is the advancement of human interactive devices. By understanding how humans process visual information, improvements can be made to interactive devices thereby enhancing the experience of users.

In trying to establish a quantitative relationship between the environment and the HVS, research has separated into two directions: i) establishing a model of HVS based on the statistics of natural images and ii) observing the responses of the HVS when stimulated by natural images. Both approaches, however, have benefited from Shannon's breakthrough in information theory [91] which established how to quantify information transmissions. To begin with, however, it is instructive to provide background to the Fourier analysis of images so that the significance of one of the main statistical regularities found in natural images is made clearer.

3.2 Background: Fourier Analysis

The properties of images can be expressed using Fourier analysis. In mathematical terms, Fourier theory provides a method to break an arbitrary periodic function into a weighted sum of sinusoidal functions. For a 1-D function $f(x)$ defined on the interval $-\pi \leq x \leq \pi$ the Fourier series is represented by:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx) \tag{3.1}$$

where a_0 is a constant, and a_n and b_n are the Fourier coefficients of f(x) given by:

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,dx \qquad \text{for } n = 0, 1, ..., \tag{3.2}$$

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\,dx \qquad \text{for } n = 0, 1, ..., \tag{3.3}$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,dx \qquad \text{for } n = 1, 2, ..., \tag{3.4}$$

Fourier analysis has aided development of harmonic analysis, which is the study of functions/signals as superpositions of waves. For example, a square-wave function, as shown in Figure 3.1 can be formed as a sum of sinusoidal components. Furthermore, it is possible to

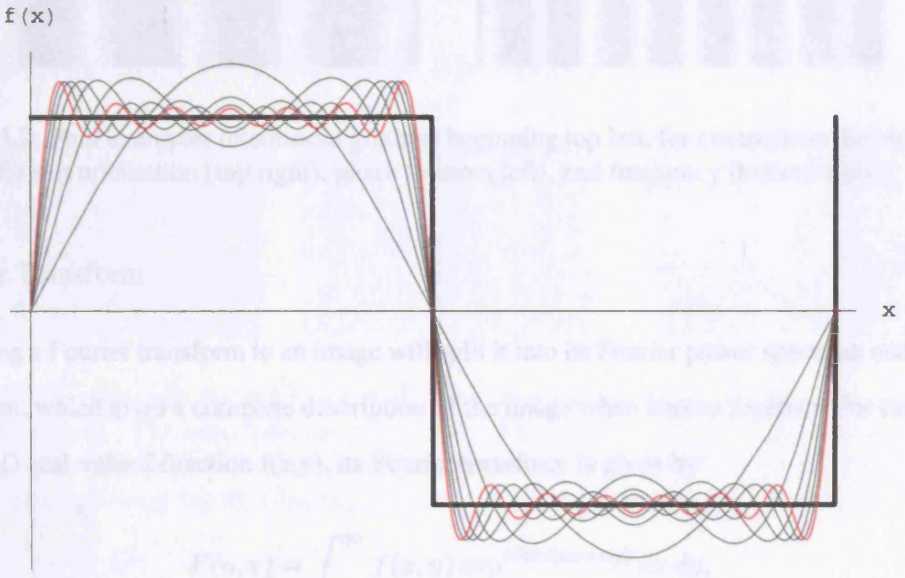


Figure 3.1: Successively summing curves of sinusoids of higher frequency approximates a square wave (bold solid black line). The red curve shows the summation of sinusoids: $\frac{1}{\pi}\left(\sin(\pi x) + \frac{1}{3}\sin(3\pi x) + \frac{1}{5}\sin(5\pi x) + \frac{1}{7}\sin(7\pi x) + \frac{1}{9}\sin(9\pi x) + \frac{1}{11}\sin(11\pi x)\right)$.

decompose a 2-D image into a summation of waves of different frequencies, for example, as sums of sinusoidal gratings. Figure 3.2 illustrates two sinusoid gratings with different orientation, phase, and frequency. Thus, an image can be represented as a 2-D function of position (ignoring scale and time) or as a signal with energy and phase. Thus, different images can be constructed with sinusoid gratings of different frequencies (i.e. the width of light and dark bands), amplitudes (luminance values of bands), phases (position of sinusoid) and orientations (angle of bars). However, it may not be obvious which sinusoid gratings would make up a particular natural scene. What is clear is that the higher frequency sinusoids will be responsible for finer detail in an image, therefore, without them, a more blurred image results.

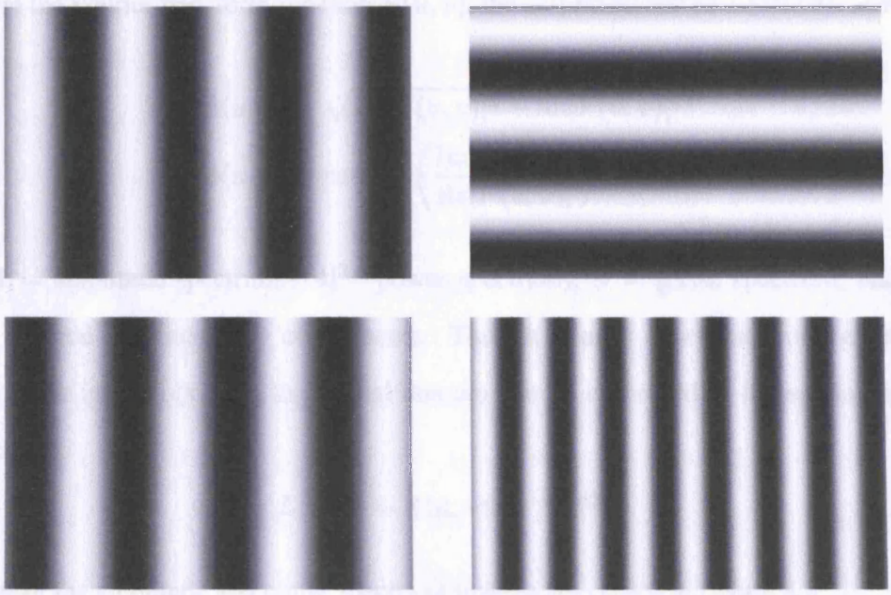


Figure 3.2: Four examples of sinusoid gratings beginning top left, for comparison are sinusoids with different orientation (top right), phase (bottom left), and frequency (bottom right).

Fourier Transform

Applying a Fourier transform to an image will split it into its Fourier power spectrum and phase spectrum, which gives a complete description of the image when known together. For example, for a 2-D real-valued function f(x,y), its Fourier transform is given by:

$$F(u,v) = \int_{-\infty}^{\infty} f(x,y) \exp^{-2\pi i(ux+vy)} \, dx \, dy, \tag{3.5}$$

and its inverse Fourier transform by:

$$f(x,y) = \int_{-\infty}^{\infty} F(u,v) \exp^{2\pi i(ux+vy)} \, du \, dv. \tag{3.6}$$

The discrete forms of Equations 3.5 and 3.6 are given by:

$$F(u,v) = \frac{1}{N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) \exp^{\frac{-2\pi i(ux+vy)}{N}}, \tag{3.7}$$

and

$$f(x,y) = \frac{1}{N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} F(u,v) \exp^{\frac{2\pi i(ux+vy)}{N}} \tag{3.8}$$

where x, y, u and v run from 0 to $N-1$.

From the Fourier transform function $F(u, v)$, the amplitude and phase can be computed:

$$A(u, v) = \sqrt{\text{Re}[F(u, v)]^2 + \text{Im}[F(u, v)]^2} \tag{3.9}$$

$$\phi(u, v) = arctan\left(\frac{\text{Im}[F(u, v)]}{\text{Re}[F(u, v)]}\right) \tag{3.10}$$

where A = amplitude spectrum ($|A|^2$= power spectrum), ϕ = phase spectrum, and Re, Im indicate the real and imaginary components. Thus, a Fourier description of the luminance function of an image L(x,y) in the spatial domain can be expressed in the spatial frequency domain as:

$$L(u, v) = A(u, v) * e^{i\phi(u,v)}. \tag{3.11}$$

Examples of the amplitude and phase spectra of images are shown in Figure 3.3.

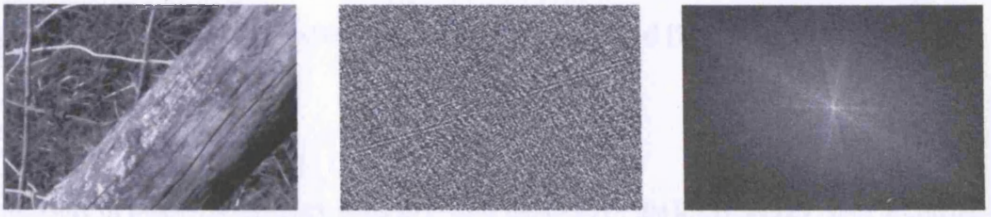


Figure 3.3: Illustrates a natural image (left) and its phase spectrum (middle) and amplitude spectrum (right). The amplitude spectrum shows that there is greater amplitude in the low frequency fourier components (light region in centre) than in the higher frequency components (darker regions toward the boundaries).

3.2.1 Local Energy

At each line and edge in an image there is phase congruency, which means that the phases of the Fourier components are equal at that particular location in the image, and when phases are congruent constructive interference occurs. This is most easily demonstrated for a 1-D step function $\epsilon(x)$ given by:

$$\epsilon(x) = \begin{cases} -1 & \text{if } x < 0, \\ +1 & \text{if } x > 0, \end{cases} \tag{3.12}$$

and which can be expressed in a Fourier representation as:

$$\epsilon(x) = \sum_{n=0}^{N-1} \frac{1}{\omega_n} \cos(\omega_n x + \frac{\pi}{2}). \tag{3.13}$$

The significant part of $\epsilon(x)$ is the behaviour at $x = 0$, where all the Fourier components are 'in phase'. At other locations $x \neq 0$, the Fourier components are 'out of phase'. This is the basis of the Local Energy Model or Phase Congruency Model [92, 93], which postulates that the HVS perceives features at locations in an image where the Fourier components are maximally in phase. In other words, where the phase information is highly ordered rather than at points of maximal intensity gradient. Furthermore, the Local Energy Model attempts to explain the results of certain psychophysical experiments on the performance of human subjects in feature tasks [94]. The model proposes that the HVS is able to process a visual signal in terms of the phase and amplitude of the signal's individual frequency components, instead of processing visual data spatially as in a Gaussian-derivative framework. However, more recent investigations suggest there is not yet a complete understanding of how the HVS approaches feature detection of edges, and therefore, neither approach to human feature detection—searching for peaks in the gradient field or peaks of local energy—should be ignored [95].

Phase Congruency

The measure of phase congruency used in [92] is the ratio of the local energy $|E(x)|$ and the sum of the amplitudes of the Fourier components A_n at the corresponding frequency. Significantly, this ratio is invariant to global changes in luminance and contrast of the image. The local energy is the local Fourier components plotted head to tail as complex vectors on an imaginary versus real axes[1]; in other words, $|E(x)|$ is the path length taken by the Fourier components from the origin to the end point of the vectors. Alternatively, $|E(x)|$ can be interpreted as $\sum_n A_n(\cos(\phi(x) - \bar{\phi(x)})$ where $\bar{\phi(x)}$ is the mean phase angle of all the Fourier components.

Having now presented Fourier analysis of images and the Local Energy Model, the characteristics of natural image statistics shall be discussed next.

3.3 Power Spectrum

In this section, the second order statistics of natural images shall be discussed. The second order statistics of natural images is presented first, specifically the ensemble mean power spectrum, because of its significance in characterising natural images.

[1]The Fourier components projected on the real axis is simply the original signal.

3.3.1 Power Law Scaling and Scale Invariance

The most prominent regularity reported in the literature is of the power spectrum of natural image ensembles. The orientation averaged power spectrum of several different natural image ensembles has been found to scale as $1/f^{\alpha}$, where α is found to be approximately 2 for an ensemble of natural images [96, 18, 97, 98, 99, 100], and f is the spatial frequency in one dimension[2]. In other words, a power spectrum of the form $1/f^2$ indicates there is equal energy in equal frequency bands (octaves). Interestingly, prior to studies on the statistics of natural images, Deriugin, aided by the results of Kretzmer on local statistics, had already found that the amplitude spectrum of television images followed a power law distribution [101, 102].

Significantly, images with power spectra of the form $S(f) \propto f^{-\alpha}$ are considered scale-invariant under renormalisation[3] meaning that computing a statistic over an ensemble of images, $\xi\{I(x)\}$ will be equivalent to computing the same statistic over the images at a different scale $I(\sigma x)$ multiplied by an appropriate constant:

$$\xi\{I(x)\} = \xi\{\sigma^{\nu} I(\sigma x)\} \ \ \forall \sigma > 0. \tag{3.14}$$

In other words, a system displaying scale invariance will have the same general statistical properties when that system undergoes a scaling transformation, therefore there is no characteristic or 'typical' scale. For example, applying the transformation $f \rightarrow af$ to a power spectrum of the form $S(f) \propto f^{-\alpha}$ gives

$$S(af) \propto (af)^{-\alpha} \propto a^{-\alpha} S(f). \tag{3.15}$$

The power spectrum can be intuitively understood in terms of pixel correlations given that the autocorrelation function and power spectrum form a Fourier transform pair (Weiner-Khinchin theorem),

$$S(f) = \int_{-\infty}^{\infty} C(x) \exp^{-2\pi i f x} \, dx \tag{3.16}$$

Thus, the power spectrum provides an alternative measure of how pixels in an image are correlated as a function of the distance between pixels.

---

[2]The power spectrum has been averaged over all orientations to produce a 1-D dependence upon spatial frequency.

[3]Mathematically, scale invariance requires that the probability density function p(x), satisfy the condition $p(\tau x) = \tau^{-\mu} p(x)$, where $\tau > 0$. It is, however, common to find this referred to as self-similarity and the condition $p(\tau x) = p(x)$ as full scale invariance.

3.3.2 Sources of Scale Invariance

Several studies attempt to determine what image contents in natural images are responsible for the $1/f^2$ scaling of its mean power spectrum. Experiments studying synthetic images have addressed whether edges, rather than occlusions, or scaling of object sizes and viewing distance, contribute to the scale-invariant property of natural image ensembles [103, 104, 105, 106]. It is unsurprising that numerous hypotheses have tried to account for the scale-invariant property of natural images. One hypothesis suggests scale invariance results from the world being made up of a collection of independent objects whose sizes and distances follow a power law distribution [98, 103, 104]. However, others have proposed that it is specifically occlusions which contribute to the power law scaling of natural images [105]. Lee and Mumford compared the statistics from a natural image database and those generated from variations of the dead leaves model (images formed from collages of independent objects which occlude one another); their study concluded that there is good correspondence—statistically—between natural and synthetically generated images. This result supports the argument that occlusions are responsible for the power law scaling observed in natural image ensembles [104].

Attempts to address all of the previously mentioned hypotheses by examining the influence of edges, occlusions and scaling through the study of simple natural image models have been made. A number of self-similar images (self-similar in terms of the sizes of objects), formed from circular disks which are occluding, non-occluding with edges, without edges, and non self-similar with an exponential size distribution of objects have been investigated [106]. The conclusions drawn from the investigation of these self-similar images is that edges do contribute to natural image ensembles having a power spectrum with power law scaling, however, self-similarity extends the spatial frequency range over which power law behaviour is found.

In summary, studies which have suggested that edges are not essential to scaling are based on a model of only the second order statistics of natural images. Results which do model the higher order statistics of an ensemble of natural images suggest that (i) edges or occlusions contribute to natural image ensembles having an average power spectrum with power law scaling, and (ii) the presence of self similar sized objects extends the power law scaling over a broader range of spatial frequencies.

3.3.3 Variation in Natural Image Power Spectra

Variations exist for the values of the power law exponent α reported in different studies for different ensembles of natural images. In the following studies [96, 18, 107, 97, 98, 99, 100], the variation of α is between 1.8 and 2.4. Individual natural images also show variation from a $1/f^2$ power spectrum. The variation between individual natural images has been reported to be between 1.5 and 3 [107, 108], which is larger than it is for the variation in the ensemble average.

Further, it has been shown that the 2-D power spectra of natural images are anisotropic over orientations; there exist horizontal and vertical orientation biases in the 2-D representation of the power spectrum [5]. These biases are in part, due to there being, on average, more horizontally and vertically oriented structures in natural images. The reported approximate $1/f^2$ form of the ensemble power spectrum of natural images is averaged over orientations. Therefore, it does not describe the ensemble power spectrum of natural images completely because of the anisotropy over orientations.

Such variations in individual natural images and the ensemble power spectrum of natural images, mean that it is important to distinguish between: i) an ensemble of synthetic images with an equivalent average power spectrum to an ensemble of natural images, and ii) an ensemble generated from maintaining the power spectrum from each individual natural image in the original ensemble. Despite both ensembles having the same ensemble averaged power spectrum, only the latter phase randomized ensemble will explicitly preserve the average power in each frequency band (orientation) of each natural image in the ensemble. The psychophysical difference between these two ensembles of images have been investigated by Tadmor and Tolhurst [109] who found that there are differences between the two classes of images, even though both ensembles are described, in other work, as the phase randomized ensemble of natural images. The influence of the power spectrum on perception, as studied in psychophysical experiments, is discussed further in Section 3.3.4.

3.3.4 Psychophysical and Physiological Significance of Power Spectra

For natural image ensembles (given that a $1/f$ scaling of their amplitude spectra is observed, which is equivalent to $1/f^2$ power spectra) investigations were designed to find out what effects, if any, occur for images with variations from $1/f$ scaling. Psychophysical experiments, by several authors were constructed to find out how well humans could perform textural discrimination tasks. Crucially, results showed that human performance is best when textures have

1/f amplitude spectra statistics [21, 22, 23, 110].

Further evidence in [111], reports that the power spectrum is important in the appearance of images. If the power spectrum of a natural image is replaced with one markedly different from the original (containing strong geometric forms), for example from a grating (see Fig. 3.2 for examples), features from the grating are visible rather than those from the original natural image. In [109], it is reported that even if the power spectrum of the natural image is altered but maintains a $1/f^{\alpha}$ form, then the image will still be degraded (although in this study the images chosen were rich in textures and shading). Moreover, even if the power spectrum is replaced with an equivalent α the image is degraded, which suggests the anisotropy in the power spectrum, that is the power in each orientation is also important in the appearance of natural images. Thus, the power spectrum of natural images is also important in determining the appearance of a natural image; the significance of phase spectra are discussed in Sections 3.5.2 and 3.5.3.

Theoretical studies, performed by van Hateren, examined the relationship of cortical neurons in the fly, and the spatiotemporal properties of natural images [112]. Given that the spatial regularities of natural images had been previously reported, it was also hypothesised in [112] that the movements unique to any particular animal provided a further regularity in the temporal visual signal of that animal. Also in the study, van Hateren proposed a theory that predicts a set of neural filters which maximise information transfer through noisy channels of limited dynamic range.

Fairly good agreement between the measured responses of the fly's visual neurons and the predictions of the model were found by van Hateren: the theory is able to produce quantitative predictions of spatiotemporal receptive fields. However, the model did produce filters that increased redundancy for low signal-to-noise ratios. This, as mentioned by van Hateren, was an inconsistency with the original hypothesis that the fly's visual neurons reduced redundancy [112]. It also highlights a point which has been ignored: what is the effect of noise in the input and throughput of the HVS. Inevitably there will be both uncertainty and variability in neural responses to identical stimuli. Thus far, in this thesis, it has been assumed that behaviour of the HVS can be predicted precisely by the input, which is not the case. The HVS is subject to internal noise and this should be factored into any complete theory describing the processes of the HVS.

Other investigations by Kardar and Zee have modelled the receptive field of barn owl neurons with a linear set of filter functions [113]. The results imply that the set of functions incorporate both audio and visual correlations of the input signal, and maximises information transfer by removing redundancies, but is also constrained by the cost of rewiring neurons (it is, from a biological perspective, more expensive to make connections between two neurons the further apart they become).

Theoretical studies by Atick and Redlich concluded that the spatial properties of mammalian retinal ganglion cells are adapted to removing redundancies caused by statistical correlations that are present in signals with an overall $1/f$ amplitude spectra, in other words, a whitening (flattening of amplitude spectrum) operation is performed on images [114]. Tolhurst and Tadmor, and Parraga and Tolhurst, created several stimuli where the amplitude spectra varies gradually from $1/f$; both their findings also support the hypothesis that the HVS is optimised to encode natural visual stimuli [115, 116]. In [116], the authors attempt to answer whether discrimination of the different amplitude spectra is achieved by i) comparing the output of channels that respond to certain frequencies, or ii) by comparing the contrast across more than one frequency? The answer reported is i) unless the images are edge enhanced, in which case the answer is ii).

However, no description has been given to how exactly the HVS could exploit the statistical regularity of ensembles of natural images having $1/f$ amplitude spectra? One possibility is performing high and low pass filtering of the image. For example, to reduce the redundancy in the image, the HVS should high pass filter the image. However, there are complications that the HVS would need to overcome. The high frequency components of an image contain a small fraction of the power in a signal and noise, which results in the high frequency components having high noise-to-signal ratios. Therefore, a balance is required between high pass and low pass filtering of the image, in order to remove redundancies as well as noise.

3.4 Contrast Statistics

Thus far, only second order statistics have been described, for example the correlations between pairs of pixels in an image. First order image statistics relate to individual pixels and not dependencies between pixels (i.e. pixels are treated independently). For example, the distribution of pixel luminance values is a first order statistic, although it has been suggested that the first

order statistics of natural image ensembles do not appear to distinguish it from other image classes [110]. Nevertheless, first order statistics—albeit along with second order statistics—have been reported to account for the low correlation between local luminance and contrast found in natural images [117]. This matches the physiological evidence that local luminance and contrast gain control mechanisms in the HVS operate independently.

Significantly, Laughlin studied the fly's visual system and found that the probability distribution of contrasts—found in the fly's natural environment—were transformed to a uniform distribution by the contrast response function (how neurons respond to stimuli of varying contrast) [118]. A uniform distribution is the maximum entropy distribution possible (for a finite range distribution); therefore, a uniform distribution implies the fly's visual system is optimised to process visual stimuli from its environment.

3.4.1 Contrast Adaptation

The contrast response function of anaesthetized cats and monkeys is not static but has been shown to change rapidly according to the mean level of contrast in a particular scene [119]. This allows the cat and monkey visual systems to be optimised to detect differences in contrast around the new mean-level-contrast. This enables adaption to the wide range of contrast levels found in their natural environment, despite individual neurons having limited range.

Adaptation to contrast has also been found in the HVS, which is unsurprising given that the HVS fixates on a new location every 200-300ms. Both slow (caused by neurons becoming less sensitive after exposure to high contrast stimuli) and fast contrast gain control mechanisms have been found in V1 [120, 121]. Furthermore, Gardner et al. demonstrated that contrast gain (the mean neuronal response divided by the stimulus contrast) changes occur in the HVS in areas V1, V2, V3 and V4. However, the response in V4 is different to the other visual cortical areas, in that a positive response is evoked whether there is an increase or decrease in contrast [122]. Gardner et al. conclude that the differences in V4 are a result of two processes occurring in the HVS. One is the contrast gain adjustments which make neurons insensitive to slow changes in contrast. This is what Gardner et al. describe as uninformative changes and a further adjustment that makes neurons sensitive to fast changes. This is important given that the correlation for either local luminance or contrast is low in going from one fixation point to another when performing 'natural tasks' [117]. This means that rapid gain control of local luminance and contrast is needed and has been found physiologically. Overall, such an

adaptation mechanism is interesting because it illustrates how the HVS is changing with respect to incoming signals, thereby transmitting the appropriate information to higher visual cortical areas.

3.4.2 Non-Gaussian Statistics of Natural Images

The power spectrum does not, by itself, reveal anything about whether or not a distribution is Gaussian. It is useful to determine whether the statistics of natural images are Gaussian because this would indicate whether higher order dependencies exist in natural images. The idea of measuring two properties—scaling and Gaussianity—at the same time was introduced by Kandanoff and is called 'coarse graining': the statistics of a coarse-grained system should remain the same despite changes in scale [123]. Ruderman observed that the scaling of the local log–contrast (local contrast is computed by taking the standard deviation of a local image patch and dividing it by the mean intensity of that region) and local gradient distributions from an ensemble of natural images is non-Gaussian [124], which implies that more than second order correlations exist.

Furthermore, Ruderman found another invariant of natural image ensembles: images formed from local pixel variances have the same statistics as those from the original image pixel intensities. This invariance was found by performing a non-linear transformation that removes the (non-Gaussian) exponential tail distribution of filter responses to the log-contrast image. Ruderman hypothesised that the exponential tails are a result of the superposition of many distributions of different variances, i.e. regions of low and high contrast. Therefore, by reducing high (low) contrast regions to lower (higher) contrast regions (normalising the log-contrast fluctuations to their standard deviation), this would remove the exponential tail distribution. For example, $\psi(x) = \frac{\phi(x)-\bar{\phi}(x)}{\sigma(x)}$ where $\phi(x) = ln(I(x)/I_0)$ is the variance modified image (I_0 is defined so that every image histogram has zero mean), $\bar{\phi}(x)$ is the local mean in a N×N block, and σ is the standard deviation of ϕ in the block.

The normalisation procedure leads to images which are noise-like with almost Gaussian statistics; this makes the information transfer optimal for neuron channels with constraints on the variance [91, 125]. It has been shown by Jaynes and Shannon that, for a continuous distribution with fixed mean, the distribution that maximises the entropy is the Exponential distribution. However if the distribution has fixed variance, a Gaussian distribution maximises entropy. Furthermore, for either continuous or discrete distributions, with finite range, a uniform distribution

maximises entropy.

3.5 Higher Order Statistics

In the main, only the first and second order statistics of natural images have been reported, however, higher order statistical properties provide important information about natural images. Essentially, the higher order statistics of images could be defined as any property of images which cannot be extracted by first or second order statistics. One example of higher order statistics is computing the correlations of more than two pixels: three pixel correlations. Petrov and Zhaoping [17] investigated the redundancy of luminance correlations in an ensemble of natural images at several spatial scales and pixel depths. It was found that the contribution to redundancy from three pixels was only 4% compared with 50% for two-pixel correlations. Petrov and Zhaoping suggested this showed that the properties of the receptive fields of V1 neurons are unlikely to have been constrained by having to measure higher than second order redundancies, however, this does not necessarily mean that other higher order correlations are not important in the HVS. Although it is interesting to note the findings made by Schneidman et al. in studying biological networks [126]. Schneidman et al. observed the responses of neurons in vertebrate retinas and show that a maximum entropy model (Ising Model) that describes only the pairwise correlations between neurons—without assuming higher order interactions—can capture over 90% of the total mutual information (i.e. including higher order correlations) between the responses of neurons. Schneidman et al., given the success of their model, suggest that despite weak pairwise correlations, many such pairs can add up to a strongly correlated network and that this can be applied to not only neuron networks but biological networks in general.

3.5.1 Gabor Filters

Field observed the histogram responses of Gabor filters applied to six natural images [18]. The Gabor representation was chosen by Field because it has properties similar to those associated with cortical cells. These properties are sensitive to certain orientations and bands of frequencies across a localised region of space [127]. Field found that the distribution of responses was markedly non-Gaussian: higher peak at zero and longer tails at extreme values compared to a Gaussian distribution, which are signs of excess kurtosis (a Gaussian distribution has zero kurtosis). Furthermore, experimental work by Baddeley et al. showed that if animals are view-

ing movies of natural scenes, the distribution of responses of visual cortical neurons is best described by an exponent function, $P(x) \propto e^{-\alpha x}$ [128]. Baddeley et al. suggested the firing rate of neurons is consistent with an optimised rate code: neurons maximise the amount of information they carry during a fixed long-term-average firing rate. However, an alternative but equivalent interpretation is provided which is that neurons minimise their average firing rate whilst keeping a fixed amount of information carrying capacity.

Field also develops a visual model [129] which utilises Gabor filters to form a representation of an image. The basic framework of the model consist of three key parts: i) *Sensors* which are the individual filters (representing complex cells), their properties include the width and length of each sensor, which is determined by the frequency bandwidth of the sensor, and the orientation bandwidth (a wider bandwidth corresponds to a shorter length); ii) *Channels* which consist of all the sensors that are sensitive to the same frequency and orientation (in two dimensions this will be an array); iii) *Code* refers to the collection of channels necessary to represent the image. At every position there are two orthogonal sensors that have phase relations in quadrature. The local energy is then the vector sum of the two sensors. However, it is the frequency selectivity that enables natural images to be represented by only a very small number of cells; this is known as sparseness.

After Field established a relationship between the statistics of natural images and the behaviour of cortical cells, more researchers attempted to produce a set of basis functions that mimic cortical simple cells [6, 130, 8, 131, 132]. In the case of Olshausen and Field, the basis functions are transformed in order to maximise the sparseness of the basis (this is achieved by having the maximal amount of bases with coefficients equal to zero) whilst retaining as much information about the image as possible. Van der Schaff and van Hateren observed the independence of complex cell responses and found that the distribution of responses was much broader (non-Gaussian with high kurtosis) for natural images than for whitened images (1/f amplitude spectra) [19]. Van der Schaff and van Hateren, therefore, concluded that cortical cells sparsely encode representations of natural images, whereas for whitened images, there is a greater probability of a maximal response in the cortical cells.

PCA and ICA

Having established that the receptive fields of simple cells are localised in space, orientation selective and bandpass, and moreover, that second order statistics (correlations) are not sufficient

to fully describe natural images (phase information cannot be ignored), it was surprising that previous work by Baddeley and Hancock attempted to decompose natural image signals into only its principle components [6]. In other words, by choosing to perform PCA[4] only linear decompositions are considered and therefore no more than second order correlations. Further, an added assumption is that natural images are stationary (invariant to spatial translations). Inevitably though, it means that principle component analysis PCA, which assumes Gaussian signals, despite natural images having non-Gaussian properties, requires additional independent component analysis ICA [110].

ICA produces a set of basis functions that are independent (although not strictly independent[133] since these models [130, 131, 19, 134] describe images as linear superpositions. In doing so, Olshausen and Field (1996) [131], and others [130, 19] have shown that when a complete population of neurons have their receptive fields optimised, in order to produce sparse representations of natural images, those receptive fields are similar to cortical simple cells (i.e. they are non-orthogonal, localised, orientation selective and band-pass, and as independent as possible). It is, however, necessary to first decorrelate the image by whitening (flattening the amplitude spectrum) before performing ICA.

Van der Schaff and van Hateren derived independent component filters from performing ICA on a large data set of natural images. Van der Schaff and van Hateren were also able to find filters which matched the receptive field properties (spatial frequency bandwidth, orientation bandwidth, length and aspect ratio) of cortical simple cells, although what is not matched is the variability in the spatial scale of the filters. The generated independent component filters are also fixed to one scale, unlike cortical simple cells, which have a range of scales.

However, ICA is not perfect in other respects. Non-linear dependencies in natural images exist across space, scale and orientation [135, 136, 137]. Therefore, beyond ICA, basis functions of wavelet transforms have been considered because such filters are sensitive to non-linear dependencies, and can therefore be used to detect whether any such dependencies exist in natural images. The choice of wavelet has been the Gabor-wavelet filter, which has been used because Gabor functions have been found to resemble simple cells in animal visual systems [138].

[4]PCA is a technique used in statistics to reduce the dimensionality of a dataset whilst retaining as much information content as possible. PCA exploits the fact that variables in the data are correlated. Usually the data is transformed to a series of weighted principle components where the first principle component will contain the most variance and successive components lower variance.

3.5.2 Phase Spectra

The power spectrum provides only a partial description of an image. Knowledge of the phase spectrum and any of its regularities in an image are interesting because phase encodes information about the localisation of edges and lines. If the phase information of a natural image is kept, but the power spectra flattened, i.e. whitened, the lines and edges of an image will be preserved [18]. Thus, edges and lines of images cannot be determined from second order statistics alone, hence, it is inferred that the skewness (third order) and kurtosis (fourth order) of image values contain phase information, as described in Equations 3.17 and 3.18.

$$\text{skew (x)} = \frac{1}{N}\frac{\sum_{i=1}^{N}(x_i - \bar{x})^3}{\left(\sum_{i=1}^{N}(x_i - \bar{x})^2\right)^{\frac{3}{2}}} \tag{3.17}$$

$$\text{kurtosis (x)} = \frac{1}{N}\frac{\sum_{i=1}^{N}(x_i - \bar{x})^4}{\left(\sum_{i=1}^{N}(x_i - \bar{x})^2\right)^{2}} - 3 \tag{3.18}$$

where $\hat{x}$ is the mean of x.

However, skewness and kurtosis are not necessarily independent of second order correlations. Therefore, in order to measure phase-only skewness and phase-only kurtosis it is necessary to first whiten the image, which removes second order correlations. By following this procedure, it has been shown that natural images have positive phase-only kurtosis (i.e. non-Gaussian distribution of image values) but natural images with randomized phases have almost zero phase-only kurtosis (Gaussian distribution of image values) [139]. This implies that there are correlations in the phase spectra of natural images.

Thomson et al. also attempt to explain what consistencies in the phase spectra of natural images might be responsible for positive phase-only kurtosis. It is argued that since kurtosis is a fourth order measurement, it must be formed from fourfold combinations of image data. Therefore, phase-only kurtosis, viewed in the Fourier domain, is defined as the integral of fourfold combinations of frequency components. Phase-only kurtosis should not be viewed in the spatial domain because the absolute Fourier phases are not invariant to translations, but all higher-order moments are invariant.

However, the space of fourfold combinations is very large, which makes it necessary to simplify the problem by considering two pairs of frequency component. From each pair, a relative phase is measured as the difference in absolute phases of its two components. The

result is two relative phase measurements from each pair. Once cast in this form it is a matter of finding if positive correlations exist between one pair of relative phases and the other pair, which would account for a positive phase-only kurtosis. Field, and later, Summers and Thomson, found positive correlation between relative spatial phases in natural images [140, 141]. The findings are supported by experimental studies by Victor and Conte who found that cortical cells are not sensitive to absolute spatial phases but are sensitive to relative spatial phases [142].

3.5.3 Psychophysical Significance of Phase Spectra

The importance of spatial phase information has been described in several psychophysical experiments: Julesz and Schumer concluded that changes to the phase spectrum of an image resulted in the image being rendered unrecognisable [143]. This is supported by Oppenheim and Lim who suggested that images are recognisable as long as the phase spectrum is adequately coded [144]. Field similarly found a dependence upon image phase information: participants would describe images with the amplitude spectra of a natural image but with randomized phases as 'unnatural' [18]. However, these results do not necessarily mean that the power spectrum of natural images makes no contribution to the appearance of images, as discussed in Section 3.3.4. In [145], it was reported that phase information was necessary to reconstruct localised features (edges and lines), in the absence of other local information. If then, by performing a local fourier transform of patches in the image, and retaining the patch location in the image, the power spectrum of the patches could be used to determine the appearance of the image.

Other authors, Piotrowski and Campbell, investigated phase perturbations: quantisation of the phase spectrum (phase values take on certain discrete values, the number of levels determines the width of bins and the total number of bins). The study found that participants could still recognise images after quantisation of the phase [146]. Further work by HüBen et al. tried to address similar questions based on images of textures. The authors found that phase quantisation beyond the 16th-level (bin widths equivalent to $2\pi/16$) in the interval $[-\pi, \pi)$ had no effect on the ability to recognise images.

Further investigations, by Thomson et al., observed the effects of phase quantisation versus phase randomization [147]. The first aspect adddressed by the authors was finding the simplest global statistics which are sensitive to image phase spectra, in contrast to observing multiple local first order statistics. Significantly, this approach was adopted not because the authors

suspected that the HVS computes higher order statistics but that it provided constraints upon how the HVS may combine responses from receptors. Thus, as mentioned previously, Thomson et al. chose to study skewness and kurtosis since non-Gaussian statistics cannot be described fully by just the mean and variance.

Thomson et al. concluded that percepts depend on the phase spectrum of individual images, not upon the type of phase perturbation used. However it is necessary to study higher order statistics, for example, the phase-only image kurtosis. Moreover, such studies use global statistical measures that pick up local sensitivities. The authors also suggest that further investigations into higher order statistics will be of benefit to studies of natural image perception, and moreover, provide further clues as to the types of relative-phase receptors which are optimised for the processing of natural images.

Experimental studies by Felsen et al. found that the responses of visual cortical complex cells (not true for simple cells which have been shown to be adapted to exploit phase redundancies for sparse coding [130, 131, 19]) had greater sensitivity to natural stimuli than random stimuli [148]. This means that visual features are more detectable from natural stimuli. However, importantly it was found that sensitivity remained high for whitened natural stimuli implying that cortical complex cells are detecting phase regularities in natural images, rather than just regularities in the power spectrum [148].

3.5.4 Geometric Properties of Natural Images

Other regularities which have been found in natural images relate to the geometrical properties of edges and lines within natural images. The simplest geometric regularity was found by observing the orientation of contours. For example there exists a prevalence of horizontal and vertical oriented contours in natural images [5, 6, 7].

The following studies detailed next, describe edge co-occurrence statistics in natural images. It should be noted that a complete comparison with the 2-point results in this thesis can not be made. This is because this thesis is concerned with the entire gradient field of images and not only where the gradient magnitude is high or where image segments belong to contours.

Nevertheless, it is possible to consider whether the 2-point results in this thesis, which measure the dependency between gradient directions as a function of both distance and angle between the two gradient measurements, are at least consistent with the edge co-occurrence statistics found by other authors; this comparison is made in Section 7.2.3.

Co-occurrence Statistics

Further investigations by Sigman et al. measured the probability of any geometric relationships which exist in natural images by observing the relationship between pairs of short image segments. What is measured is the probability of finding pairs of short segments which have a certain position relative to each other and then recording the orientations of those segments. How the probability changes with distance between the segment pairs and angle between the pairs is considered. Sigman et al. find that the most probable arrangement for segments adjacent to each other (side-by-side) is to be iso-oriented and the least probable is to be perpendicular. When two image segments are iso-oriented then the most probable spatial arrangement is to be part of a common line (co-linear).

Iso-orientation is significant because iso-orienation is a property which is exploited in the HVS; the response of cells in V1 display side-inhibition [149, 150, 151], which occurs when a segment lies inside the cell's classical receptive field as well as another segment (orthogonal to the receptive field orientation) just outside the cell's classical receptive field (non classical receptive field). Further, for pairs of segments which have different relative orientations then a co-circularity rule is obeyed, that is the segments are maximally correlated when arranged on a common circle.

Sigman et al. conclude that there exist strong long-range correlations for collinear segments in natural scenes. Sigman et al. also suggest that these long range correlations extend the concept that the output of linear-oriented filtering of natural scenes cannot be statistically independent. In summary, Sigman et al. found that the most significant geometric structures in natural visual scenes are the line and the circle [9].

A similar study by Geisler et al. also observed the co-occurrence statistics of image segments as a function of the distance, orientation difference between segments, and direction of second element with respect to the first. Edges are extracted from images by using Log Gabor functions, which resemble the responses of cortical neurons in V1, to perform a two-stage filtering process. Images are first filtered with bandpass filters and with bandpass and oriented selective Log Gabor filters. What is interesting here is that a distinction is made as to whether or not the pairs of edge elements are on the same contour; something which was not differentiated in [9]. Nevertheless, Geisler et al. find that nearby edge segments that are either co-linear or co-circular with the reference edge segment have greater probabilities of occurring in natural

images which is consistent with [9].

Chow et al. in [10] extended the work Sigman et al. by investigating what structures could produce the observations made in [9] . The hypothesis made in [9] is that natural scenes contain objects of varied sizes and the continuity and smoothness of those object boundaries lead to the finding of a co-circularity rule. Thus, to test the hypothesis, Chow et al. generated synthetic images from simple geometric objects: stadia and ellipses. Chow et al. concluded that even a small number of nearly circular objects—ellipses and stadia (two semi-circles joined by lines)—as well as circles would produce statistics consistent with a co-circularity rule. Furthermore, it is possible to extend the results to include segments of closed contours and that the presence of only a few circular arcs would also contribute to co-circularity. Therefore, the results of Sigman et al. show that it is likely natural scenes have many closed smooth contours, not just circles.

Lee and Kardar [152] observed the properties of lines in natural images by decomposing the power spectrum of natural images into transverse and longitudinal Fourier components. These two components measure the variation in power along orthogonal directions—parallel and perpendicular—relative to a given wavevector in Fourier space. It was found that the Fourier spectrum of natural images exhibited more power in their transverse component than longitudinal component [152]. Lee and Kardar concluded that natural images must have an abundance of extended sharp edges; moreover, filters that remove much of the redundancy in such images were also constructed. Lee and Kardar were able to show some qualitative consistency with their theoretically constructed filter function, and the lateral connections (connections which link columns of neurons with similar orientation preference, OP) found between co-linear and co-oriented neurons in V1. The consistency found was that the constructed filters were strongest for stimuli that are co-linear with parallel OPs, or co-circular for two neurons with orthogonal OPs (a consistent result with [9]) and, therefore, the optimal connection between two V1 neurons is dependent on their orientation preference and correlations in the input signal.

Edge Statistics and Cortical Complex Cells

In another paper, van der Schaff and van Hateren, observed the statistics of edges in an ensemble of natural images. Each image was transformed so that edges of a particular orientation can be extracted [8]. This kind of transformation resembles the function cortical complex cells perform, therefore, the study was expected to illuminate some of the coding properties of cortical complex cells. Previously, it had been proposed that sparse coding in cortical complex

cells occurred due to the existence of edges in natural images in comparison to synthetic images (Gaussian noise images). Van der Schaff and van Hateren wanted to address whether the responses of complex cells are independent and sparse.

Van der Schaff and van Hateren found that in natural images there are extended spatial dependencies between pairs of edges if the second edge is along the preferred orientation (along the initial edge direction) and albeit weaker correlation in directions other than the preferred orientation. Synthetic images dropped to zero correlation beyond the effective size of the receptive fields, moreover, at the same separation, the correlation in natural images was found to be small (< 0.5 bits). Dependency is also found between responses at the same location but with filters that have different orientation preferences. However, such dependency drops rapidly after effectively one orientation bandwidth of a cortical complex cell. Furthermore, dependency across frequency scale drops rapidly after scale factors larger than 2.8 (this factor is equivalent to a frequency bandwidth ~ 1.5 octaves) to 0.3 bits.

What is slightly paradoxical is that previously it has been stressed how important sparse representations are to encoding images efficiently. However, van der Schaff and van Hateren have showed that much of the dependency in natural images is attributable to edges. Given sparsity increases with independent responses to stimuli, it is odd that the HVS should have neurons that are orientation selective. Van der Schaff and van Hateren address this paradox by asking whether it is possible to form a representation of natural images using independent filters which are not orientation selective.

To examine this, an isotropic model of local variance (similar to local contrast but compares logarithmic not linear intensity values) in natural images is studied using 7×7 pixel patches with an effective size equivalent to the representation of cortical receptive fields. Therefore, a comparison of the mutual information shared between two local variances and that of two responses from cortical cells can be made.

Van der Schaff and van Hateren found that, for natural images, the dependency amongst local variances is greater than for orientation selective filters [8]. Therefore, orientation selective responses are more independent than local variances for natural images. However, synthetic images showed no change in dependencies between spatially separated orientation selective responses and local variance responses. Furthermore, the responses showed independence even in natural images once the filters are spatially separated by more than the effective extent of the

receptive field [8].

However, the observation that filters at different scales can be independent, once spatially separated beyond the extent of the receptive field, is in contrast to the single-scale independent component filters generated by ICA. A possible explanation for this, given by the authors, is that a filter with a large receptive field may overlap many smaller receptive fields. Thus, it may be possible, by observing all the small receptive fields, to determine what the response of the large receptive field will be, thus making them dependent despite the output of each small filter being independent.

3.6 Summary of Natural Image Statistics

In this chapter a number of important regularities in natural images have been highlighted in order to motivate and support the work in this thesis. Much of the literature emphasises the scale-invariant property of natural images from the observation that the average power spectrum of natural image ensembles follows a power law decay ($1/f^2$) with respect to spatial frequency (averaged over orientation). The power spectrum is essentially a second order statistic, but there exist higher order statistical regularities, i.e. correlations between more than pairs of pixels in an image. Another important spectrum of images is the phase spectrum, which has been reported to also have a power law decay. This implies that phase regularities exist in the phase-only kurtosis and phase-only skewness distributions of natural images.

The significance of both power spectra and phase spectra in the appearance of natural images have also been studied. Investigations suggest that both contribute to the appearance of natural images. The ensemble of power spectrum of natural images, which is averaged over orientations, is also not necessarily sufficient to describe features in individual natural images (in particular natural images rich in textures and shading, or containing strong geometric forms) because the power spectrum can be biased in particular orientations. The phase spectrum has been shown to contribute to features such as lines and edges in natural images, but without the original power spectrum, significant degradation in the image such as blurring can occur, and loss of features at certain orientations.

Geometric regularities have also been reported such as the co-circularity rule: given two short image segments in natural scenes the most likely contour joining them is either a line or circular arc. Experiments investigating the cause of such a regularity conclude that natural

images are composed of many closed smooth contours and objects of different sizes.

However, apart from highlighting the statistical regularities of natural images, a continuing theme that has been developed throughout is the relationship between natural image statistics and the evolution of the human visual system (HVS). Evidence has been presented to support the efficient coding hypothesis: the HVS is optimised to encode natural visual stimuli. It has been reported that even as early as the retina, ganglion cells are performing a whitening operation (flattening of the power spectrum) on visual stimuli, which reduces redundancy in the signal.

Furthermore, neurophysiological experiments have shown visual cortical neuron activity is suppressed when the visual system is being subjected to natural visual stimuli compared to random stimuli. Such experiments have been performed on not only human subjects but also in the fly, cat and monkey, to show that sparse coding is a necessity for other biological organism with visual systems. Essentially, therefore, sensory neurons are limited in how much and how quickly they can transmit data, which affects the evolution of any living organism's visual system.

Psychophysical experiments have also shown observers are better at discriminating synthetic images with the same power spectrum power law decay as natural image ensembles. Furthermore, computational experiments have derived filter functions based on efficient coding constraints: filters which reduce redundancy, maximise information transmissions, and increase independency between filters begin to resemble cortical visual receptive fields.

In conclusion, the benefits of investigating how the HVS has evolved to encode natural image stimuli is not just restricted to furthering understanding of the HVS, it is also beneficial for developing computational algorithms to solve perceptual tasks. Thus, the study of natural image statistics is worthwhile and motivates the main work of this thesis.

Chapter 4

Technical Aspects to Computations and Methodology

This thesis is concerned with the statistics of gradient directions in natural images; in particular statistics that describe how gradient directions at different locations are coupled, which includes how the coupling changes with separation. The coupling is quantified using measures of dependency rather than correlation because it is possible for two variables to be uncorrelated but not independent if there are non-linear dependencies between the variables. However, the reverse is true, independence between two variables means the correlation is zero. For example, if one variable X depends on the square of the other Y^2, the correlation will be zero but there is certainly a dependence between X and Y. Thus, in order to evaluate the amount of dependency between gradient measurements, information-theoretic methods, in particular interaction information (higher order mutual information), which was first introduced by McGill [153], later by Fano [154] and in more formal detail by Han [155], have been used. Moreover, interaction information is particularly useful for the treatment of more than two variables. Interaction information is discussed in Section 4.1.

In practice, to obtain an estimate of the gradient direction dependencies, a joint probability distribution of gradient directions is formed. To estimate Shannon's measure [91] of entropy (Sec. 4.1.1), entropy estimators are applied to the joint probability distributions. Each entropy estimator has a particular bias and variance. Therefore, a selection of entropy estimators are discussed in Section 4.2 to reduce the bias, and in Section 4.3, the bootstrap procedure is discussed as a method to assess the variance of the entropy estimator. Lastly, in Section 4.4, an introduction to the application of derivative Gaussian kernels to images is presented.

4.1 Information Theory

Information theory is an appropriate framework for establishing whether exclusive 3-point interactions exist (synergetic dependencies between three measurements); by definition an interaction is structure that cannot be described by any subset, i.e. any 2-point dependencies. For example, given two variables X and Y, which may or may not be dependent on each other, then if knowledge of a third variable Z is acquired, the dependency between X and Y can change in the context of Z. Let us then suppose that X and Y are independent, knowledge of Z can then result in X and Y becoming dependent. The reverse is also possible if X and Y were originally dependent; knowledge of Z can result in X and Y becoming independent. Furthermore, knowledge of Z may not change the dependency between X and Y. Moreover, degrees between the possible outcomes mentioned exist and interaction information provides a method of quantifying such degrees. In other work [156], if a third attribute increases the dependency (mutual information) between two other attributes it is referred to as synergy. Conversely, if the dependency decreases, this is referred to as redundancy. Both synergy and redundancy are described in Section 4.1.3. However, because information theory can be attributed to the breakthrough made by Shannon on information entropy, this is discussed first.

4.1.1 Shannon's Measure of Entropy

Shannon [91] introduced information entropy, as opposed to thermodynamic entropy (a measure of disorder in a system). Shannon provided theories on the communication of discrete and continuous data (and a mixture of both), however, the focus here shall be on the discrete description as this is most relevant for the type of analysis used in this thesis. Information entropy is a measure of uncertainty in a given probability distribution (information entropy). The measure of information entropy H proposed by Shannon, for events i, is:

$$H = -\sum_i p_i \log p_i \tag{4.1}$$

It turns out this is the only choice for H that can satisfy the following conditions:

1 H should be in continuous in p_i, i.e. a small change in p_i does not cause large changes in H.

2 $H = 0$ if the outcome is certain, i.e. if and only if $p_i = 0, \forall i$ except one.

3 H is a maximum when all events are equally likely, i.e. $p_i = 1/n, \forall i$.

An alternative interpretation of H is to consider the 'surprise' of an event occurring, which is also known as self-information: how surprised or informed are we when a certain event occurs. An initial guess is that for a given event a, which has a probability of occurring of p_a, our surprise s_a of this event occurring is equal to $1/p_a$. However, if an event is composed of two independent events, with the probability of occurring p_b and p_c (i.e. $p_a = p_b.p_c$), the total surprise should be the sum of these events: $s_a = s_b + s_c$, but this gives $1/p_b + 1/p_c$, which is not equal to $1/p_a$. Instead, the definition for the measure of surprise for an event x, with probability p_x, should be $s_x = -\log p_x$, which gives a total surprise of $-\log p_a = -\log(p_b.p_c) = -\log p_b - \log p_c$. Thus the entropy is given by the surprise s_x, weighted by the probability of the event p_x, therefore entropy is expected surprise.

The measure of entropy in Equation 4.1 is easily extended to joint events X and Y. There may be several outcomes for each event, so let $p(x_{i=I}, y_{j=J})$ be the probability of outcome $x_{i=I}$ together with outcome $y_{j=J}$. The joint entropy is then given by:

$$H(X,Y) = -\sum_{i,j} p(x_i, y_j) \log p(x_i, y_j) \tag{4.2}$$

and the marginal entropies are given by:

$$H(X) = -\sum_{i,j} p(x_i, y_j) \sum_{j} \log p(x_i, y_j) \tag{4.3}$$

$$H(Y) = -\sum_{i,j} p(x_i, y_j) \sum_{i} \log p(x_i, y_j) \tag{4.4}$$

$H(X,Y)$ is then a measure of how uncertain we are about the pair of random variables X and Y. It is also possible to construct a measure of conditional entropy, for example $H(X|Y)$, which describes how the entropy of X decreases with knowledge of Y:

$$H(X|Y) = -\sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(y_j)}, \tag{4.5}$$

where, $p(x_i) = \sum_j p(x_i, y_j)$, $p(y_j) = \sum_i p(x_i, y_j)$, and $p(x_i|y_j) = \frac{p(x_i,y_j)}{p(y_j)}$. In other words $H(X|Y)$ is a measure of how much uncertainty remains about the value of X after knowing Y.

4.1.2 Mutual Information

Using the joint entropy and marginal entropies, it is possible to construct Shannon's measure of mutual information:

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \tag{4.6}$$

$$= -\sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}. \tag{4.7}$$

I(X;Y) is the amount of uncertainty removed from X with knowledge of Y, or it can be interpreted the other way: how much uncertainty remains in Y given X. This is because $I(X;Y)$ is a symmetric measure, meaning that $I(X;Y) = I(Y;X)$. Furthermore, there is self-information $I(X;X)$, which reduces to $H(X)$.

Mutual information can also be written in terms of relative entropy. This leads to the more intuitive interpretation that mutual information measures the 'distance' between the joint probability distribution $p(x_i, y_j)$ and the product of their marginal distributions $p(x_i) \times p(y_j)$. The measure of the distance between probability distributions used for this is known as the Kullbeck-Leibler (KL) divergence. For the probability distributions $p(x)$ and $p(y)$, the KL divergence is defined as:

$$KL(p(x)||p(y)) := \sum_i p(x) \log \frac{p(x_i)}{p(y_i)}, \tag{4.8}$$

which is i) always positive $KL \geq 0$, and ii) equal to zero only when $p(x) = p(y)$. Thus, rewriting mutual information in terms of KL divergence:

$$I(X;Y) = KL(p(x,y)||p(x)p(y)). \tag{4.9}$$

Mutual information can also be written in terms of conditional entropies:

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X). \tag{4.10}$$

4.1.3 Interaction Information: Synergy and Redundancy

A useful depiction of interaction information is through Venn diagrams, which has been shown to be relevant by Yeung [157] because of the similarity between set theory and information-

theoretic operations such as exclusion and inclusion. This also provides an alternative description of interaction information, in a set theoretic sense, of mutual information as an intersection of entropy, joint entropy as a union, and the conditional entropy as a set difference, but this interpretation requires the association of abstract sets in place of random variables, therefore the formal description in terms of set theory shall not be given here. Furthermore, the area of regions in the Venn diagram are not necessarily proportional to the amount of uncertainty, especially in cases where the interaction information can be negative; such problems are discussed in [158].

Nevertheless, Venn diagrams help to visualise the terms involved in interaction information. The most familiar example of interaction information is that between two variables, which has been described in Equation 4.6 as mutual information and is illustrated in Figure 4.1. Figure 4.1 shows terms like $H(X)$, which denote the entropy of X, and $H(X|Y)$, which is the entropy of X conditional on Y. It also shows the mutual information shared between X and Y, which is denoted $I(X;Y)$.

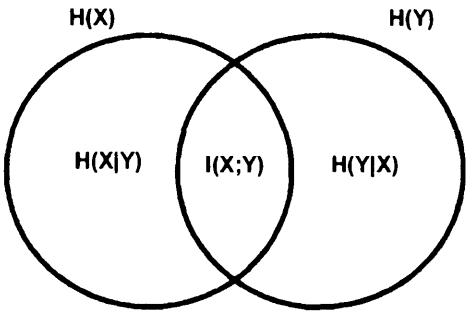


Figure 4.1: Illustrates a venn diagram of the mutual information $I(X;Y)$ between two random variables X and Y.

If mutual information is extended to three variables the Venn diagram in Figure 4.2 is obtained. The important area of Figure 4.2 is the region which overlaps in the centre $I(X;Y;Z)$. This is the mutual information between X, Y, Z. In order to begin to interpret $I(X;Y;Z)$ compare Figures 4.1 and 4.2. One observes that the region of $I(X;Y)$ (Fig. 4.1) is replaced by the sum of $I(X;Y;Z)$ + $I(X;Y|Z)$ in Figure 4.2, and therefore:

$$I(X;Y;Z) = I(X;Y) - I(X;Y|Z), \tag{4.11}$$

or expressed in terms of joint and marginal entropies:

$$\begin{aligned} I(X;Y;Z) =& H(X) + H(Y) + H(Z) \\ & - (H(X,Y) + H(X,Z) + H(Y,Z)) + H(X,Y,Z). \end{aligned} \tag{4.12}$$

The quantity $I(X;Y|Z)$ is the conditional mutual information between X and Y in the context of Z (i.e. when there is knowledge of Z). In terms of joint entropies, $I(X;Y|Z)$ is given by:

$$\begin{aligned} I(X;Y|Z) =& H(X,Z) + H(Y,Z) - H(Z) - H(X,Y,Z) \\ =& H(X|Z) - H(X|Y,Z). \end{aligned} \tag{4.13}$$

From the joint and marginal entropy terms, the interpretation of $I(X;Y|Z)$ can be read as the reduced uncertainty in X due to knowledge of Y in the context of Z. The interpretation of $I(X;Y;Z)$ (the interaction information of X, Y, Z) is the difference between the mutual information of X and Y, and the conditional mutual information of X and Y (i.e. given Z), which can be stated more succinctly as the reduction in mutual information between X and Y when Z is known. Equivalently, an alternative interpretation of $I(X;Y;Z)$ is how much uncertainty is removed from Z by jointly knowing X and Y which is given by $I(X,Y;Z)$, compared to the amount of uncertainty removed from Z due to knowledge of X which is given by $I(X;Z)$, and the uncertainty removed from Z due to knowledge of Y which is given by $I(Y;Z)$. Therefore, we can rewrite $I(X;Y;Z)$ as

$$I(X;Y;Z) = I(X;Z) + I(Y;Z) - I(X,Y;Z). \tag{4.14}$$

If $I(X;Y|Z) = 0$, it follows from Equation 4.11 that $I(X;Y;Z)$ must be nonnegative, further, if $I(X;Y;Z)$ is positive then $I(X;Y)$ must also be positive. Also, if $I(X;Y;Z) = 0$, then $I(X;Y|Z) = I(X;Y)$, this means no triplewise dependencies exist. This is because of one of two reasons: i) only pairwise dependencies exist between A and B, i.e. $p(A,B,C) = p(A,B)p(C)$, or ii) all three variables are independent of each other i.e. $p(A,B,C) = p(A)p(B)p(C)$. Therefore, if the dependency between X and Y increases when informed by Z, compared to without knowledge of Z, then $I(X;Y;Z)$ is negative, and when this occurs there is said to be synergy between X, Y, and Z. If the opposite occurs

where the dependency between X and Y decreases when informed by Z, i.e. $I(X;Y;Z)$ is positive, then there is said to be redundancy between X, Y, and Z. It is also noteworthy that $I(X;Y;Z)$ is symmetric such that all permutations of X, Y, and Z are equivalent, e.g. $I(X;Y;Z) = I(Y;Z;X) = I(Z;Y;X)$.

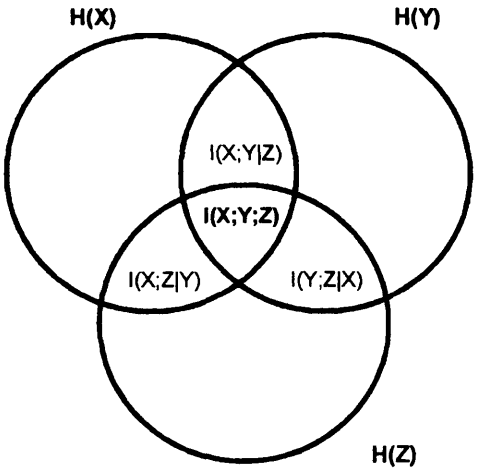


Figure 4.2: Illustrates a venn diagram of the interaction information $I(X;Y;Z)$ and conditional mutual information $I(X;Y|Z), I(X;Z|Y), I(Y;Z|X)$ between three random variables X, Y and Z.

An alternative viewpoint is to express mutual information in terms of relative entropy by following Equation 4.2:

$$I(X;Y;Z) = KL(p(x,y,z) \parallel q(x,y,z)) \tag{4.15}$$

where $q(x,y,z) = \frac{p(x,y)p(x,z)p(y,z)}{p(x)p(y)p(z)}$.

Therefore, the interaction information $I(X;Y;Z)$ is the KL distance between the probability distribution of $p(x,y,z)$ and $q(x,y,z)$. However, the behaviour of this KL divergence is not an appropriate 'distance' measure because $q(x,y,z)$ does not necessarily satisfy the normalisation condition of a true probability distribution: $\sum p = 1$, which can lead to $KL < 0$. Therefore, it is more common to define the total ternary mutual information, which is given by:

$$I(X,Y,Z) = KL\left(p(x,y,z)\right) \parallel p(x)p(y)p(z)\right), \tag{4.16}$$

and in terms of joint and marginal entropies by:

$$I(X,Y,Z) = H(X) + H(Y) + H(Z) - H(X,Y,Z). \tag{4.17}$$

This is also referred to as the total correlation [159]. $I(X,Y,Z)$ measures the total dependence between X, Y, and Z and therefore is always nonnegative, and zero if and only if the three variables are independent of each other, i.e. $p(A,B,C) = p(A)p(B)p(C)$. Even if $I(X,Y,Z)$ is positive, it does not necessarily imply there are triplewise dependencies. For example, if $p(A,B,C) = p(A,B)p(C)$ there are pairwise dependencies between A and B but no triplewise ones.

4.2 Entropy Estimators

Thus far, the entropy calculations have been based on knowing the true probability distribution p. In practice, only a sample of this distribution is available by making a finite number of observations N and placing them into M states. Therefore, only an estimate of p and the entropy H can be made. This leads to both statistical and systematic errors for such entropy calculations. For example, if p_i, in Equation 4.1, is replaced with $\hat{p}_i = \frac{n_i}{N}$, where n_i are the number of occurrences of event i, the naive entropy estimator (maximum likelihood estimate) is produced:

$$\hat{H} = -\sum_{i}^{M} \hat{p}_i \log \hat{p}_i. \tag{4.18}$$

The naive estimator results in a systematic underestimation (bias) of the entropy H due to statistical fluctuations that make the distribution look less uniform. In other words, the expectation value of $\hat{H}$, which is the estimate of the entropy, is different from the true value of the entropy H, i.e. $E(\hat{H} - H) \neq 0$, where E denotes the expectation value. The bias exists because entropy is a non-linear function of n_i. Furthermore, because the logarithm is a concave function, the bias is negative, which can be shown using the Jensen inequality[1]. It is also clear that as $\hat{p} \to 0$ unusually high scores for the entropy are obtained because $\log \hat{p} \to \infty$. To reduce this bias, correction terms are added to the naive estimator. However, there is a trade-off between reducing the bias and increasing the statistical error.

To estimate the error, a Taylor expansion of the logarithm term in Equation 4.18 can be

[1]The Jensen inequality is $E[f(x)] \geq f(E[x])$ for convex functions, and $E[f(x)] \leq f(E[x])$ for concave functions.

applied; expressing Equation 4.18 in terms of n_i and N:

$$\log N - \frac{1}{N}\sum_{i=1}^{M} n_i \varphi(n_i). \tag{4.19}$$

The naive estimator is then given by the first order term $\varphi(x) = \log x$, and the second order correction term is $-1/2x$, as proposed by Miller [160]. Thus, the Miller entropy estimator is given by:

$$\log N - \frac{1}{N}\sum_{i=1}^{M} n_i(\log(n_i) - 1/2n_i). \tag{4.20}$$

Other estimators that reduce the bias, which are in fact asymptotically unbiased for large N, at the cost of increased variance have been proposed by Grassberger [161], one of which is:

$$\varphi(x) = \Psi(x) - \frac{(-1)^x}{x(x+1)} \tag{4.21}$$

where $\Psi(x) = \frac{d}{dx}\ln\Gamma(x)$, and $\Gamma(x) = (x-1)!$. However, the estimators proposed by Grassberger are normally applied in situations where $p_i << 1$, i.e. where there are insufficient samples taken compared to the number of states; this results in several states having zero frequency.

In Figure 4.3 is an illustration of the effect of Miller's correction on the entropy score estimated from two different histogram distributions. In the first case, Figure 4.3 left, has a low number of counts n in each bin and the difference in entropy score between the naive and Miller entropy estimators is of the order of 10^{-1} (nats). In the second case, Figure 4.3 right, which has a higher number of counts in each bin, the difference between the two estimators is of the order of 10^{-2} (nats).

4.3 Variance of Entropy Estimators: Bootstrap

Not only is there an associated bias with any entropy estimator but there is also a certain amount of variance. We have chosen to calculate this variance using the bootstrap procedure introduced by Efron [162]; it is a method that can be used to obtain statistical inference about both the bias or variance of a given estimator $\hat{\Theta}$ of a statistic Θ. It is based on a resampling procedure which assumes that your sample is a good representation of the true population. By randomly resampling (i.e. each element of the original sample has an equal chance of being sampled) with replacement from the original sample (not the true population) one computes a number of

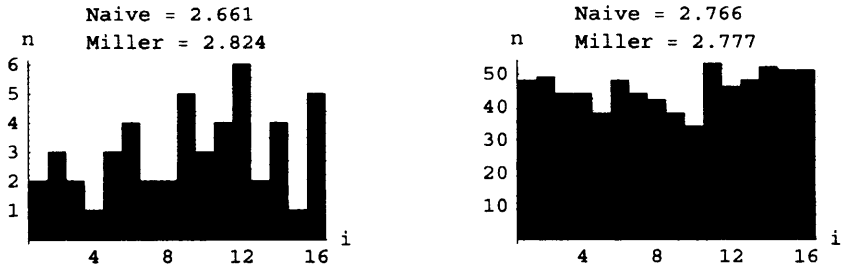


Figure 4.3: The entropy scores for both histograms are computed using the naive estimator and with Miller's correction, which are quoted above each histogram. On the left the bins i are populated with a low number of counts n in each bin. This results in a significant difference between the naive entropy estimator and Miller's correction. In contrast, the histogram on the right has an adequate number of counts in each bin and hence the difference between the estimators is smaller.

'bootstrap samples'. Each bootstrap sample has the same number of elements as the original sample but may contain duplicate elements. The number of bootstrap samples that should be taken is determined empirically. Formally the number of bootstrap samples should be infinite, however, the bootstrap procedure converges quickly with a greater number of samples and therefore, depending on the precision required, a small but finite number of samples can be taken. There will also be practical limitations such as the computation time as to how many samples can be reasonably gathered.

The new scores are then computed by applying the same estimator $\hat{\Theta}$ to the 'bootstrap samples', which leads to a new distribution of $\hat{\Theta^{(b)}}$ that differs from the original sampled distribution. The bootstrap distribution is an estimate of the sampling distribution of the true distribution. Thus to estimate the pc-percentile confidence limit of $\hat{\Theta}$ we first calculate the standard deviation S of the bootstrap distribution of $\hat{\Theta}^{(b)}$ by:

$$S(\hat{\Theta^{(b)}}) = \sqrt{< (\hat{\Theta}^{(b)})^2 > - < \hat{\Theta}^{(b)} >^2} \tag{4.22}$$

and then multiply $S(\hat{\Theta^{(b)}})$ by $\sqrt{2}\,\mathrm{erf}^{-1}\,\rho$.

The key point is that the bootstrap method, unlike other statistical inference methods, does not make any a priori assumptions as to what shape this distribution takes (although it is possible to do a 'parametric bootstrap' by sampling from a known distribution). Therefore, the bootstrap method relies heavily on the assumption that the sample distribution is a good estimate of the true population and increasing the number of 'bootstrap samples' increases the accuracy of the bootstrap estimates.

However, for the purposes of the analysis in this thesis, and which is unique to this work, it has been necessary to modify the bootstrap procedure to deal with sampling from histograms which contain the frequency of data rather than raw data itself. To summarise, the steps involved in this non-parametric bootstrap procedure from histogram counts are:

1. The data are a collection of histograms T_x where $x = 1, \ldots, M$, which form the sample of the true population.

2. Randomly draw with replacement a histogram $T_{x=X}$.

3. Re-populate each bin in $T_{x=X}$ by randomly resampling from a Poisson distribution $P(z) = \frac{e^{-\lambda}\lambda^z}{z!}$ with mean λ equal to the original bin count[2], where z is the bin label (event). Treating each bin count as a random Poisson variable is appropriate for estimating the statistical fluctuations that occur within a given interval/bin due to the discrete property of bin counts. In other words, we want to estimate the bin count again if we repeat the sampling; the bin count in this repeated sample is independent of the original sample.

4. Repeat steps 2 and 3, M number of times, to create M samples: $T_x^{(b)}$.

5. Sum all the bootstrap sample histograms together to form one total bootstrap histogram, $T^{B'} = \sum_{x=1}^{N} T_x^{(b)}$

6. Compute the statistic of interest Θ (in this thesis, this will be the entropy of the distribution) using the same estimator $\hat{\Theta}$, i.e. $\hat{\Theta}(T^{B'})$.

7. Repeat the above steps, N number of times, to create N number of bootstrap estimates for the statistic of interest $\hat{\Theta}_{y=1,\ldots,N}(T^{B'})$.

Lastly, to calculate the ρ-percentile confidence limits of the statistic of interest we compute the standard deviation S of $\hat{\Theta}_{y=1,\ldots,N}(T^{B'})$ multiplied by $\sqrt{2}\,\mathrm{erf}^{-1}\,\rho$.

If we follow the non-parametric bootstrap on sampled histogram data which have been sampled from three different distributions: Normal, Laplace and Uniform; we would expect the confidence limits on the entropy estimates to reflect the different uncertainty in estimating the entropy. For comparison, Figure 4.4 left column shows an example of three different sample histograms together with their bootstrap samples to their right. As expected the entropy

[2] Although the total number of counts in each bootstrap sample may be different from $T_{x=X}$, the total number of bootstrap sample histograms is still M, as per step 4.

estimates for the sampled histograms taken from Laplace and Uniform distributions are below and above that of the Normal distribution; moreover, the confidence limits, estimated from the bootstrap procedure, are also above and below that of the Normal distribution.

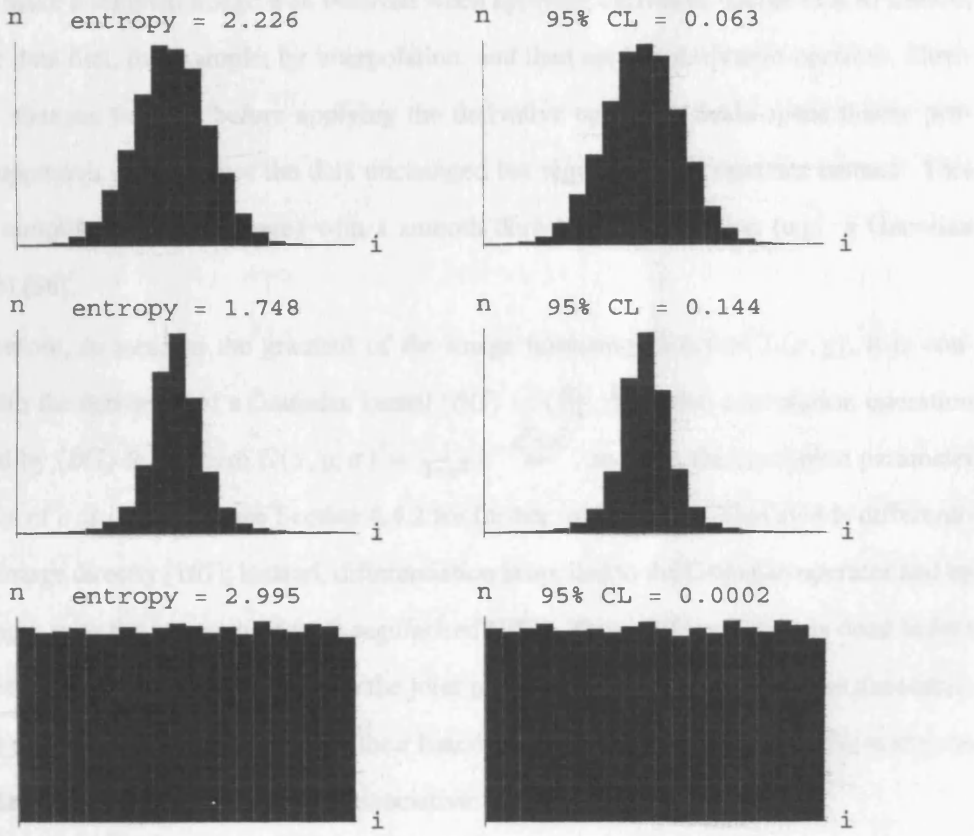


Figure 4.4: Left column: top shows one example of a sample histogram taken from a Normal distribution with its entropy estimate quoted, middle from a Laplace distribution, and bottom from a Uniform Distribution. To the right of each sample histogram in the left column is one example of its bootstrap sample, with the 95% confidence limit (CL) in the entropy estimate quoted. The labels n and i on the vertical and horizontal axes indicate the number of counts and bins respectively.

4.4 Applying Gaussian Derivatives to Images

Differentiation of a discrete image is not well defined. A technique to solve this problem is known as regularisation [163]. More generally, regularisation is a technique applied to data to make it behave well (this means that small changes in the input to a function leads to a small change in the function's output, as outlined by Hadamard [164]) when certain operators (e.g. differential operators) are applied to that data. In the context of image analysis, regularisation forms the cornerstone of (Linear) Scale-Space theory which was first proposed by Iijima [165]

and later, independently, by Witkin [166], and Koenderink [79].

4.4.1 Linear Scale-Space Theory

A step to make a sampled image well behaved when applying derivative operators is to smooth the image data first, for example, by interpolation, and then apply a derivative operator. However, this changes the data before applying the derivative operator. Scale-space theory provides an approach which leaves the data unchanged but regularises the operator instead. This involves sampling the data (image) with a smooth derivative test function (e.g. a Gaussian derivative) [86].

Therefore, to measure the gradient of the image luminance function $L(x, y)$, it is convolved with the derivative of a Gaussian kernel $(\partial G) = \{\frac{\partial G}{\partial x}, \frac{\partial G}{\partial y}\}$; this convolution operation is denoted by $(\partial G) \otimes L$ where $G(x, y; \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$, and σ is the resolution parameter (the choice of σ is important, see Section 4.4.2 for further information). This avoids differentiating the image directly [167]; instead, differentiation is applied to the Gaussian operator and by convolving it with the image the data is regularised [167]. Thus, differentiation is done before any smoothing of the data. This relies on the joint properties of convolution and the derivative, which are both linear operators and so is their functional composition. Therefore, the operation of convolution and taking derivatives is associative:

$$\partial(G \otimes L) = G \otimes (\partial L) = (\partial G) \otimes L \tag{4.23}$$

The choice of using a derivative Gaussian kernel is not arbitrary. Firstly, Gaussian derivative functions have been used as basis functions to model the receptive fields of simple V1 cells in primates [168, 169, 170, 171]. Secondly, the foundation of scale-space theory emphasises that vision is about making physical observations, therefore, observations are made through non-zero sized apertures. The Gaussian kernel happens to satisfy a number of constraints upon the properties of this aperture function when it acts as an operator on an image. In a perfect world (mathematics) this aperture could be made infinitesimally small thereby it would have a resolution of zero. However, in reality, such an aperture would capture no light; therefore, the aperture should be of non-zero size. Furthermore, the aperture should be uncommitted, for example there should be no preferences in the image for: i) location (shift invariance), ii) orientation (isotropy) of structures, and iii) size of structures (scale invariance).

The constraints on the aperture have consequences on its form and application: it is appropriate to convolve the kernel with the image (this also regularises the image), the kernel integrates over a circular area, and successive observations of the image lead to a wider kernel; the Gaussian kernel satisfies these properties. It is also possible to derive the Gaussian kernel based on arguments of causality: this means that by blurring an image no extra structure ('spurious detail') should arise [79]. This assumption, along with invariance to translation and rotation, and treating all scales equally, leads to certain constraints on the relationship between first order derivative changes in scale ($\frac{\partial u}{\partial t}$) and the second order derivative change in space ($\frac{\partial^2 u}{\partial x^2}$). In 1-Dimension, this means that as $\frac{\partial^2 u}{\partial x^2}$ decreases (< 0), $\frac{\partial u}{\partial t}$ increases (> 0) and vice versa, so that $\frac{\partial^2 u}{\partial x^2}\frac{\partial u}{\partial t} > 0$. In other words, in moving to coarser scales, local maxima in the gradient field always decrease and local minima increase.

This leads to the familiar form of the diffusion equation $\frac{\partial^2 u}{\partial x^2} = \alpha\frac{\partial u}{\partial t} > 0$ of which the Gaussian kernel is the Green's function (i.e. it is the solution to this inhomogeneous differential equation subject to boundary conditions: as $t \rightarrow 0$, the original unblurred image is obtained). The quantity being diffused is the intensity of the image over scale σ ($t = \sqrt{2}\sigma$). Interestingly, the Gaussian kernel can also be derived by considering the entropy of the observed signal [172]. The main concept in this derivation is that an uncommitted observation maximises entropy under the constraints previously mentioned; in other words, the disorder is greatest when the observation has no preference for certain structure.

4.4.2 Errors Gaussian Derivative Operators

The accuracy in calculating the gradient of a slope from sampled data using the derivatives of Gaussian kernels is affected by both the order of differentiation and the scale of the kernel [173]. For a given derivative order n, the value of the derivative worsens for smaller scales σ. This is most easily understood in the Fourier domain as an aliasing effect caused by band limitation. Gaussian derivatives act as bandpass filters, as shown in Figure 4.5. This can be shown analytically, by considering the Fourier transform of Gaussian derivatives:

$$F\{\frac{\partial^n}{\partial x^n}G(\sigma,x)\} = \int_{-\infty}^{\infty} G(\sigma,x)e^{-2\pi i\omega x}dx \qquad (4.24)$$
$$= (-i\omega)^n F\{G(\sigma,x)\}.$$

where $G(\sigma,x) = \frac{1}{\sqrt{\pi}\sigma}e^{-x^2/2\sigma^2}$.

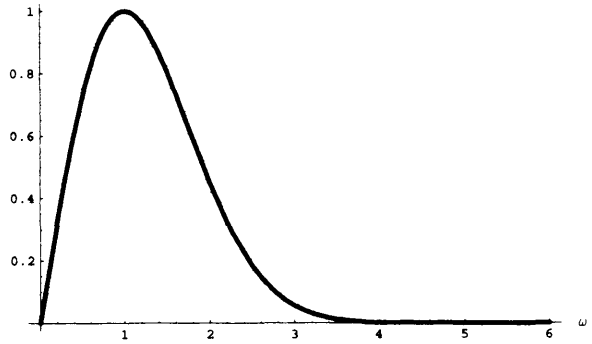


Figure 4.5: Illustrates the first order Gaussian derivative kernel normalised power spectrum. The maximum of the curve occurs at $\omega = 1$; generally for n-th order derivatives the maximum occurs at $\sqrt{n}$.

For instance, the Fourier transform of the Gaussian kernel is defined from $-\pi$ to π. As the scale of the kernel is decreased in the spatial domain, i.e. its extent decreases, this consequently broadens the kernel in the spatial frequency domain. Essentially as σ decreases in the spatial domain, more and more signal energy lies outside the kernel's defined range in the spatial frequency domain. Pictorially this is seen as an inability to fit the zero crossings of the Gaussian derivatives inside their Gaussian envelopes (higher order derivative equals higher number of zero crossings). Mathematically, the error is expressed as the amount of power which leaks from the kernel (i.e. in the aliased frequencies) relative to the the total power:

$$\text{error}(n, \sigma) = \frac{\int_{\pi}^{\infty} \omega^{2n} e^{-\sigma\omega^2/2} d\omega}{\int_{0}^{\infty} \omega^{2n} e^{-\sigma\omega^2/2} d\omega} \tag{4.25}$$

For lower order derivative Gaussian kernels this error is small; it is necessary to go up to order 4 before the error rises above 0.01. Thus, in practice, if a sufficiently large scale is chosen the aliasing error becomes insignificant. Therefore, for the computations in this thesis, a scale of 4 is chosen (unless otherwise stated) when applying first order derivative Gaussian kernels to measure the gradient.

Another type of error results from the amplification of low amplitude but high frequency noise by Gaussian derivatives. This has been studied by Blom et al. [174] who suggest that if the scale of the Gaussian derivative is sufficiently large the error is not significant. The amplification is greater for higher order derivatives, but we can apply first order Gaussian derivatives of scale 4 without such problems.

Chapter 5

Methodology: Data Collection

In this chapter several methodologies are presented which include: how the four main image classes are generated: natural, phase randomized natural, whitened natural and Gaussian (Sec. 5.1); how the gradient directions are measured (Sec. 5.2); how the histogram data is collected (Sec. 5.3 and Sec. 5.4); and how the mutual information scores (Sec. 5.5) and their confidence limits (Sec. 5.6) are calculated.

5.1 Generating Image Classes

For the natural image class, a set of 100 images are selected from a subset, which excludes images suffering from motion blur and saturation [175], of the van Hateren database [8]. The images are 1024 by 1536, 12-bit images of outdoor scenes; examples of which are shown in Figure 5.1.

Phase randomized natural images are produced by randomizing the phases of the natural images. First, the Fourier transform of an image I is computed. Second, the phases of the Fourier components are set to random values. Third, an inverse Fourier transform is computed to return to the spatial domain.

In contrast to phase randomized natural images, each whitened natural image is generated by leaving the phase of a natural image unaltered but setting the power of each Fourier component to unity. Finally, the class of Gaussian noise images consist of images with an image function whose luminance values are drawn from a Normal distribution. The Fourier transform of Gaussian noise images has a flat power spectrum with random phases. For comparison, an example from each image class is given in Figure 5.2.

Studying this quadruple of image classes will add further understanding as to the importance of the power and phase spectra of natural images and the amount of its gradient direction

Figure 5.1: Four examples of natural images taken from the van Hateren database [19]. Images contain a mixture of sky, vegetation and man-made structures.

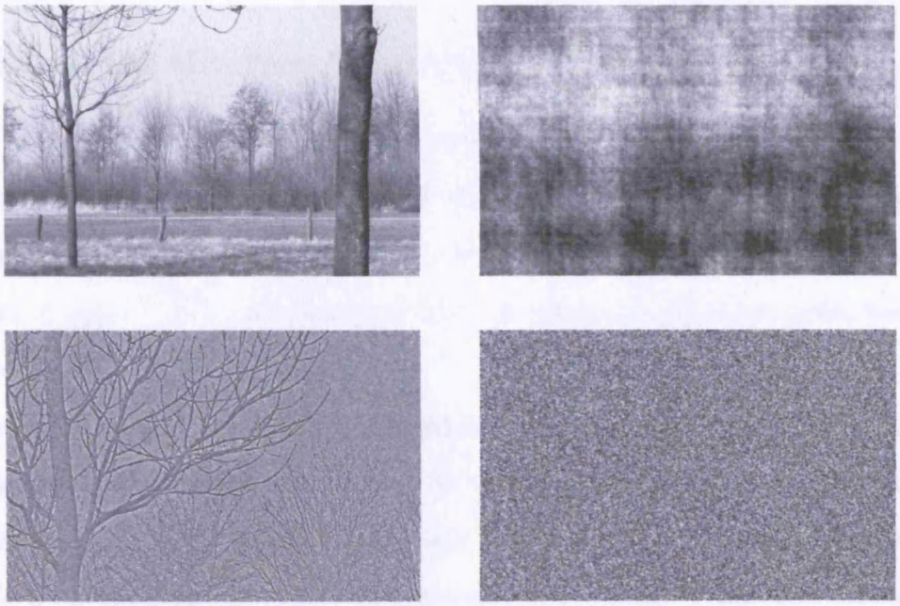


Figure 5.2: Examples from the four image classes used in this study. From top to bottom and left to right: natural image, phase randomized natural image, whitened natural image (zoomed in), and Gaussian noise image (zoomed in).

dependencies by grouping results in terms of similar power spectra or similar phase spectra. This grouping is possible because phase randomized natural images contain only the power spectrum information of natural images, and whitened natural images contain only the phase information from natural images. Furthermore, Gaussian noise images have neither the power nor phase spectra information of natural images, but have flat power and random phase spectra. A representation of the image classes' similarities and differences is illustrated in Figure 5.3.

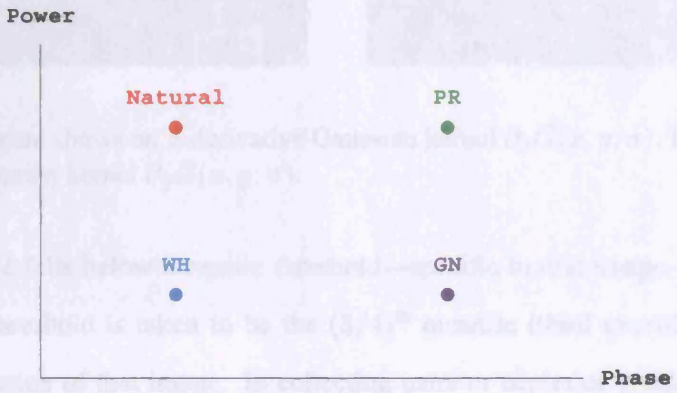


Figure 5.3: A representation of the similarities between the image classes in a power spectrum vs. phase spectrum space. Image Key: natural (Natural), phase randomized natural (PR), whitened natural (WH), and Gaussian noise (GN).

5.2 Extracting Gradient Directions and Magnitude

In order to measure first order derivatives of an image (as explained in Section 4.4.1) its luminance function $L = L(x, y)$ is convolved with the partial derivatives of a Gaussian kernel ∂G (see, Fig. 5.4 for examples of the partial first order derivatives of a Gaussian): $(\partial G) \otimes L$ where $G(x, y; \sigma) \equiv \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$ and σ is the resolution parameter ($\sigma = 4$ is used in this work unless otherwise stated).

The gradient direction at a point in the blurred image luminance function $L(x, y; \sigma)$ is then found by taking the arctangent of the ratio of the partial Gaussian cartesian first derivatives. To avoid boundary effects, it is necessary to window the image after convolution: points which are within 7σ ($>$ extent of kernel) of the image boundary are excluded. Alternative methods could be implemented at the boundary such as extending the boundary by reflection, wrapping around to the opposite boundary, or truncating the kernel.

Furthermore, for some experiments gradient directions where the gradient magnitude is large in each image are observed, i.e. gradient directions occurring at locations where the

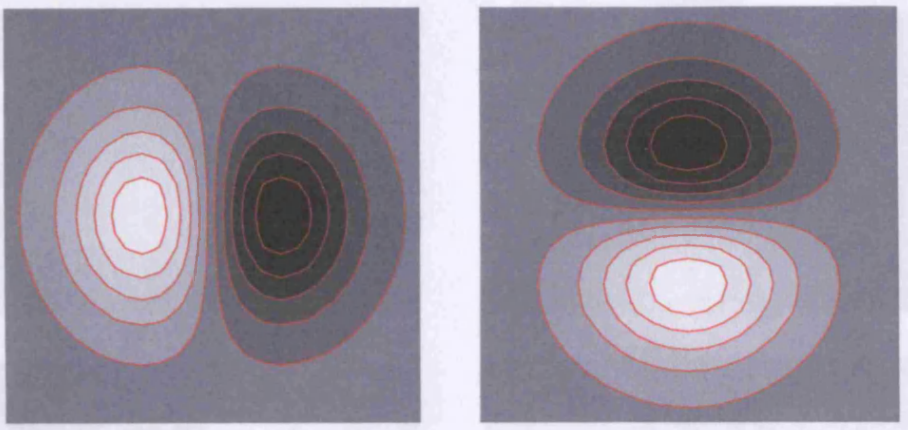


Figure 5.4: Left figure shows an x-derivative Gaussian kernel $\partial_x G(x, y; \sigma)$. Right figure shows a y-derivative Gaussian kernel $\partial_y G(x, y; \sigma)$.

gradient magnitude falls below a certain threshold—specific to that image—are ignored. For each image, the threshold is taken to be the $(3/4)^{\text{th}}$ quantile (third quartile) of the gradient magnitude distribution of that image. In collecting pairs or triples of gradient directions, the threshold must be satisfied by all the gradient directions.

An example of a natural image and the gradient directions where the gradient magnitude is large are shown in Figure 5.5. Further, in Figure 5.6, the x- and y-derivative of the natural image in Figure 5.5 left are shown, together with the gradient direction and magnitude.

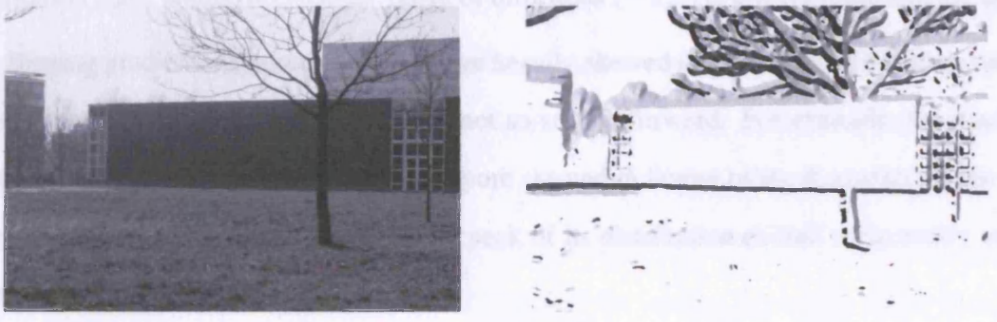


Figure 5.5: Left figure shows a natural image and right is the gradient direction where the gradient magnitude is high. The gradient directions are encoded as a grey scale value. The scale of filter σ used is 4.

5.3 Data Collection

Data is collected by computing the values of the gradient direction and the gradient magnitude for an ensemble of images (from the same image class) and forming histograms. In the case of gradient directions the data is binned into 32 equal-sized bins of width $\pi/8$ for the 1-point data

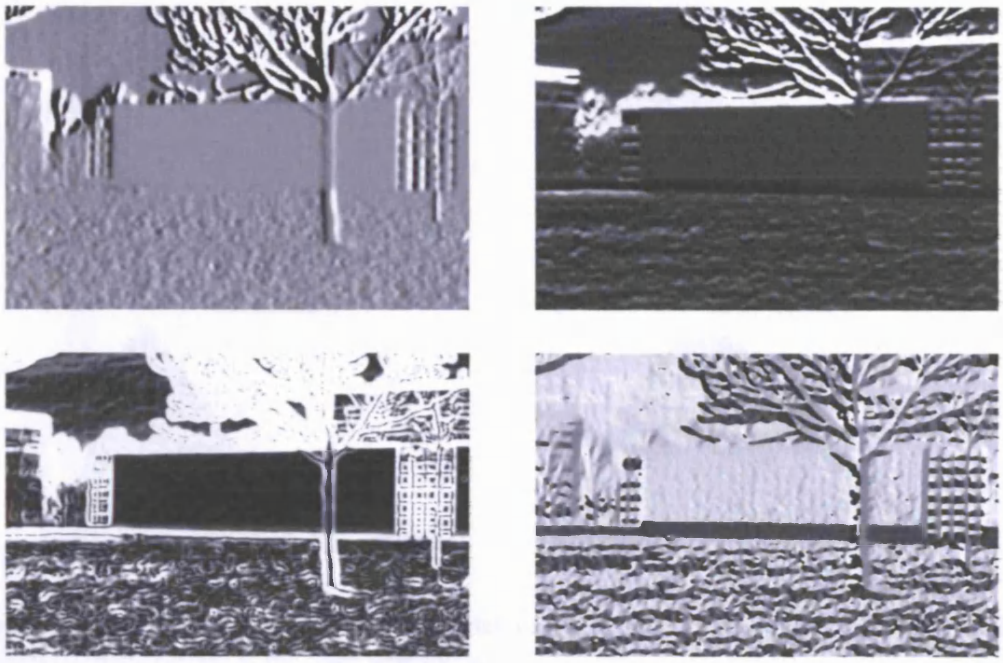


Figure 5.6: Top row: the x- and y- derivative images of Fig. 5.5 are shown left and right respectively. Bottom row: gradient magnitude and gradient direction image. The scale of filter σ used is 4.

and 16 for the 2-point and 3-point data. This choice for the number of bins allows a sufficient number of counts in each bin to be collected in a reasonable time. Binning the data for gradient directions is made simple because the range of directions $(-\pi,\pi]$ is always finite and the same.

Binning gradient magnitudes, which have heavily skewed distributions of differing ranges for each image class (shown in Fig. 5.7) is not so straightforward. For example, the gradient magnitude distribution of natural images is more skewed in favour of weak gradients than it is for phase randomized images which has the peak of its distribution shifted horizontally to the right toward stronger gradient magnitudes.

The n^{th} bin boundary has been set at the $(n/16)^{\text{th}}$ quantile value of each individual image giving a total of 16 bins per dimension. The gradient magnitude histograms are formed for each image separately and summed together to form a total histogram for all images. The quantile values are determined from 100,000 gradient magnitudes in each image.

For 1-point statistics, one million gradient measurements are collected from a set of 100 images (10,000 gradients from each image) from the same image class. The one million gradient measurements from each image class are binned as described in the previous section to form two 1-D histograms of gradient directions and gradient directions where the gradient magnitude

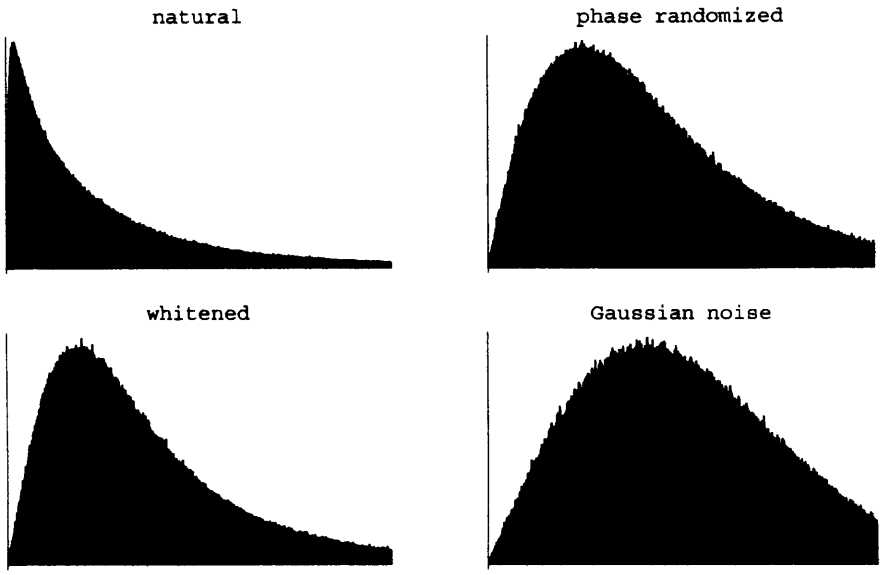


Figure 5.7: Histograms of gradient magnitudes in the four different image classes. The upper range of each histogram is the 95th percentile.

is large for each of the four image classes.

For the 3-point measurements, there are several choices over the spatial configuration of each location. For simplicity, a randomly oriented equilateral triangular configuration (see Fig. 5.8) has been chosen. From all the 3-point configurations possible, the equilateral triangle is the only one which is symmetrical and where all three points are equidistant from each other, therefore, it is a natural configuration to examine first.

The gradient at each vertex of a triangle is recorded for 10,000 triangles per image for 100 images. The gradient direction is measured relative to the line drawn from the triangle's centre of symmetry to the respective locations at each vertex. This is repeated for different separations ($k = 0.25, 0.5, 0.75, 1, 1.5, 2, 2.5, 3, 3.5, 4, 5$) between the vertices of the triangle, in multiples of the filter scale σ. This enables the study of how the 2-point and 3-point gradient direction dependencies change with separation between the measurements in a manner independent of the scale of the Gaussian derivative kernel.

For each individual image and value of k, the 3-point statistics are binned to form three 3-D histograms. Each 3-D histogram represents the gradient magnitude, gradient direction and strong (where the gradient magnitude is large) gradient direction statistics, denoted by $s = 1, 2, 3$ respectively. Thus, for each image class and value of k there are 100 histograms $u_{s;t=1,2,\ldots,100;k}$ which are summed to produce a total histogram: $U_{s;k}(A, B, C)$.

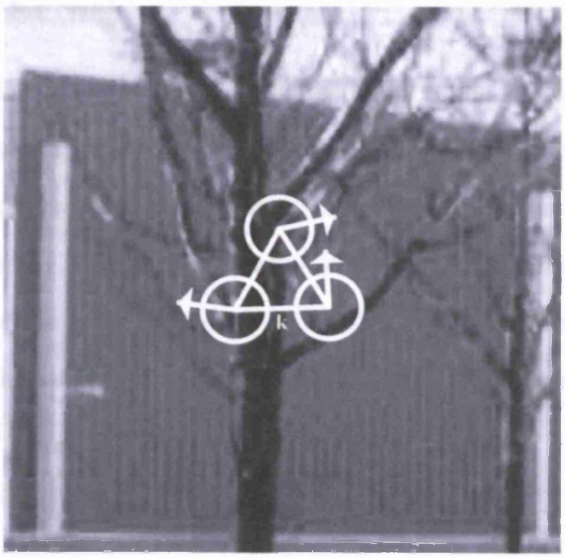


Figure 5.8: Illustrates the triangular spatial configuration adopted in the 3-point measurement (triples of gradient measurements). White arrows illustrate gradient directions. Circles have radius $\sqrt{3}\sigma$, which is the effective extent of the first derivative Gaussian kernel and the length of triangle sides are 5σ, i.e. $k = 5$.

For a given $U_{s;k}(A, B, C)$, three 2-D joint histograms can be produced by collapsing one dimension: $U_{s;k}(A, B)$, $U_{s;k}(A, C)$, $U_{s;k}(B, C)$, which represent the results for 2-point statistics. Furthermore, collapsing two dimensions produces three 1-D marginal histograms: $U_{s;k}(A), U_{s;k}(B), U_{s;k}(C)$.

One example of a joint and marginal histogram for gradient directions is shown in Figure 5.9. Note that a total of one million gradient triples is sufficient so that the mean bin count in a histogram is not close to zero, which makes the estimation of information-theoretic scores less accurate. For the 3-D histograms, which are 16^3 in size, the mean bin count is 244.

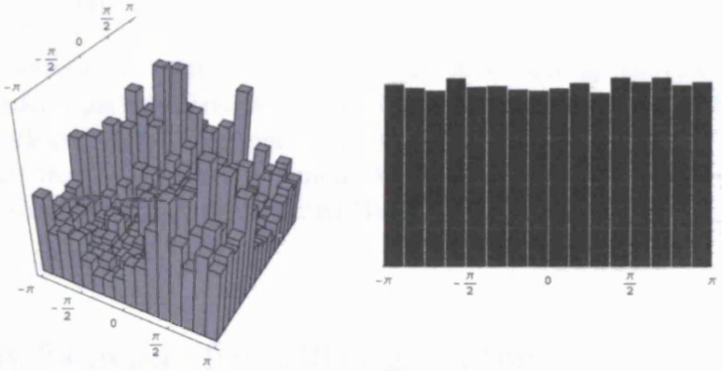


Figure 5.9: Illustrates one example of a 2-D histogram (left) formed from the 2-point statistics of gradient directions, and one of its 1-D marginal histograms (right).

5.4 Overlap of Kernels

A consideration in understanding the gradient direction dependencies is the physical extent of the operators used to calculate them. If the operators overlap heavily, it would be expected that there are some dependencies between the measurements; overlap does occur when the distance $k\sigma$ between measurements is small. How much overlap occurs is illustrated in Figure 5.10. Note that the effective extent of the first derivative Gaussian operator can be approximated analytically from the position of the zero crossing of the partial Gaussian cartesian third derivative: $\frac{\partial^3}{\partial x^3} G(x;\sigma) = \frac{x^3 - 3x\sigma^2}{\sigma^6} G(x;\sigma)$. Thus, the effective extent is $\sqrt{3}\sigma$ from its centre, which is illustrated in Figure 5.11.

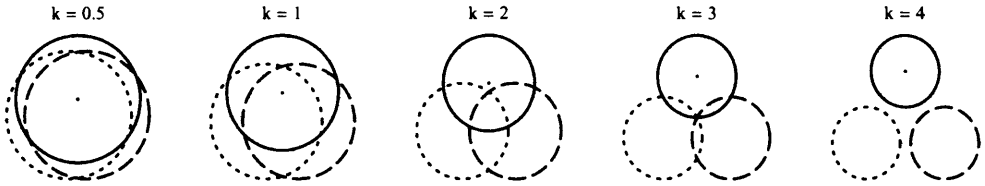


Figure 5.10: Illustrates the effective overlap of the triangular spatial configuration of first order derivative Gaussian kernels ($\sigma = 4$) as shown in Fig. 5.8 but for smaller k values $\{0.5, 1, 2, 3, 4\}$. The radius of each circle is $\sqrt{3}\sigma$: the effective extent (see, Fig. 5.11) of the kernels.

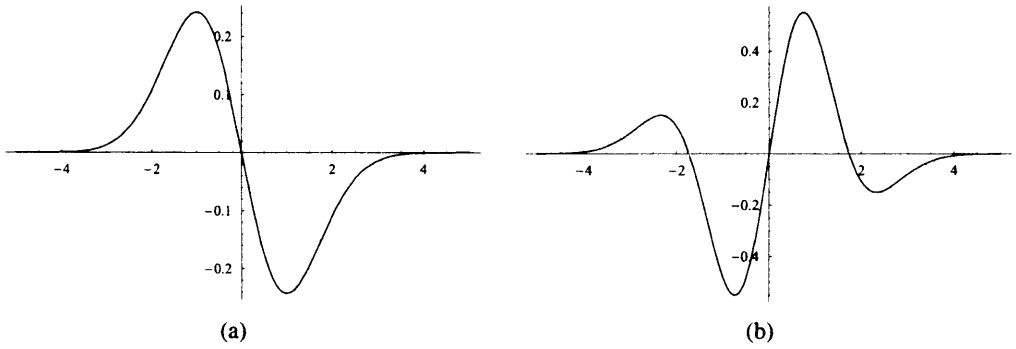


Figure 5.11: Illustrates how the effective extent of the first order derivative Gaussian is calculated. The first order x partial derivative in one dimension is shown in (a). The distance from its centre to its inflexion points is denoted by the grey vertical lines in (a). The distance of the inflexion points from the centre are computed from the first zero crossing (marked by grey vertical lines in (b)) of the third order x partial Gaussian derivative (illustrated in (b)).

5.5 Entropy Estimates from Histogram Data

To calculate the mutual information between pairs of gradient magnitudes or directions at a given separation, estimates are made of the entropy of the 2-D joint histogram of gradient

magnitudes or directions $\hat{H}[U_{s;k}(A,B)]$, and the entropy of its two marginal distributions, $\hat{H}[U_{s;k}(A)]$, $\hat{H}[U_{s;k}(B)]$.

The maximum likelihood (naive) estimator with Miller's correction [160] is chosen to reduce the bias of the naive estimator but reduce the variance of other estimators with higher order corrections:

$$\hat{H} = \ln N - \frac{1}{N}\sum_{i=1}^{M} n_i(\ln n_i - \frac{1}{2n_i}) \tag{5.1}$$

where N and M are the total number of counts and bins respectively, n_i is the number of counts in bin i. Once the joint and marginal entropies are known the mutual information is calculated from:

$$I_{s;k}(A;B) = \hat{H}[U_{s;k}(A)] + \hat{H}[U_{s;k}(B)] - \hat{H}[U_{s;k}(A,B)]. \tag{5.2}$$

Given that the entropy scores in Equation 5.1 are measured using natural logarithms, the units of information are nats (for bits, logarithm base 2 is used). For the 3-point information, we calculate the following interaction information using the 3-D histogram:

$$\begin{aligned} I_{s;k}(A;B;C) =& \hat{H}[U_{s;k}(A)] + \hat{H}[U_{s;k}(B)] + \hat{H}[U_{s;k}(C)] + \hat{H}[U_{s;k}(A,B,C)] \\ & - \left(\hat{H}[U_{s;k}(A,B)] + \hat{H}[U_{s;k}(A,C)] + \hat{H}[U_{s;k}(B,C)]\right). \end{aligned} \tag{5.3}$$

5.6 Bootstraps

Here a description is given as to how the bootstrap samples are formed from the histogram data in order to compute the confidence limits on the information-theoretic calculations. To create a bootstrap sample, 100 (for each image class there are 100 images in each ensemble) histograms are drawn randomly with replacement from $u_{s;t;k}$. For each sample histogram $u_{s;t=T;k}$ chosen, the bins are repopulated for that histogram by treating each bin count as a Poisson variable to form 100 new histograms $u'_{s;t;k}$. Treating the bin counts as Poisson variables simulates the statistical fluctuations caused by discrete bin counts. A new total bootstrap histogram $U^{(b)}_{s;k}(A,B,C)$ is then formed by summing $u'_{s;t;k}$. From $U^{(b)}_{s;k}(A,B,C)$, the joint and marginal distributions can be formed, as shown for $U_{s;k}(A,B,C)$. This process is repeated 200 times. To calculate the 2-point and 3-point mutual information confidence interval the bootstrap scores

are computed for $r = 1, 2, ..., 200$ by (for brevity we have omitted the subscripts of $U^{(b)}_{r;s;k}$):

$$I^{(b)}_{r;s;k}(A;B) = \hat{H}[U^{(b)}(A)] + \hat{H}[U^{(b)}(B)] - \hat{H}[U^{(b)}(A,B)] \tag{5.4}$$

$$\begin{aligned} I^{(b)}_{r;s;k}(A;B;C) =& \hat{H}[U^{(b)}(A)] + \hat{H}[U^{(b)}(B)] + \hat{H}[U^{(b)}(C)] + \hat{H}[U^{(b)}(A,B,C)] \\ & - \left(\hat{H}[U^{(b)}(A,B)] + \hat{H}[U^{(b)}(A,C)] + \hat{H}[U^{(b)}(B,C)]\right). \end{aligned} \tag{5.5}$$

Chapter 6

Results

In this chapter, the 1-point histogram distributions (Sec. 6.1), 2-point dependency (Sec. 6.2) and 3-point dependency (Sec. 6.4) of gradient directions, gradient magnitudes (2-point and 3-point only) and gradient directions where the gradient magnitude is high (strong gradient directions) for ensembles of the four image classes: natural, phase randomized natural, whitened natural and Gaussian noise are presented. Also, 2-point gradient direction dependencies for individual natural images are presented (Sec.6.3).

In this chapter, the gradients have been measured using first order derivative Gaussian kernels at a scale of $\sigma = 4$ in all cases. Results at different scales are not shown because it is found that gradient direction dependencies do not change at different scales for the four image classes: natural, phase randomized natural, whitened natural, and Gaussian noise, examined in this chapter.

6.1 1-point Statistics

Figure 6.1 shows the histogram of gradient directions obtained from natural images, phase randomized natural images, whitened natural images and Gaussian noise images.

Figure 6.1 shows that natural images have a slight excess of vertically (peaks at $\uparrow, \downarrow$) and horizontally (peaks at $\rightarrow, \leftarrow$) oriented gradients, which is consistent with the statistics of oriented contours [5, 6, 7]. The fact that for natural images there are more vertically up oriented gradient directions than down is not surprising given that natural images in the van Hateren set contain sky (lighter) and ground (darker); sunlight and skylight in the natural environment come from above. This is supported by the illustration in Figure 6.2 of the average luminance image of 100 natural images from the van Hateren set.

It is less well-known that phase randomized natural images have a prevalence of vertically

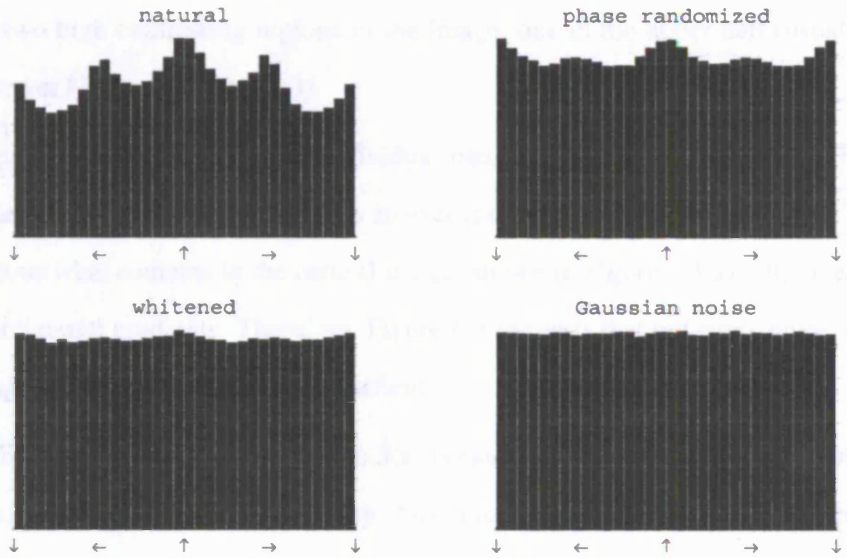


Figure 6.1: Histograms of gradient directions for the four different image classes. Along the horizontal axis tick marks $\leftarrow, \rightarrow$ indicate horizontally oriented gradients and tick marks at $\uparrow$ and $\downarrow$ indicate vertically upward and downward oriented gradients. Natural images show a prevalence of horizontally and vertically oriented gradients (top left); phase randomized images show a prevalence of horizontally oriented gradients only (top right); whitened natural images (bottom right) and Gaussian noise images (bottom right) have no orientation preference for gradients.

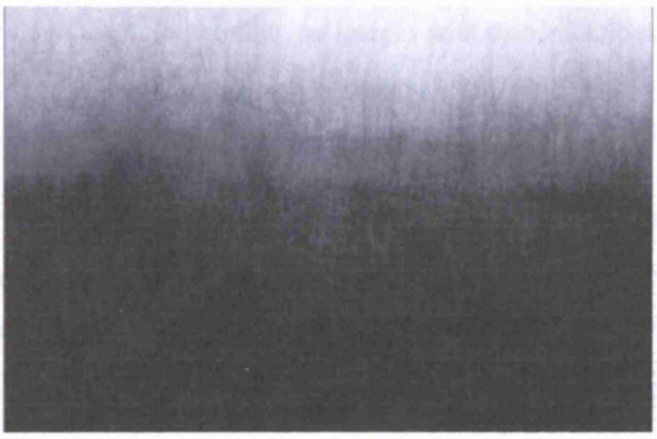


Figure 6.2: Illustrates the average of 100 van Hateren natural images. It shows a light (due to sky) region in the upper part of the image and a darker (due to ground) region in the lower part.

oriented gradients. Investigations reveal that such a prevalence is again due to the majority of the natural images (from which the phase randomized versions are computed) in the ensemble containing two high contrasting regions in the image, one in the upper half (usually sky) and one in the lower half (usually ground).

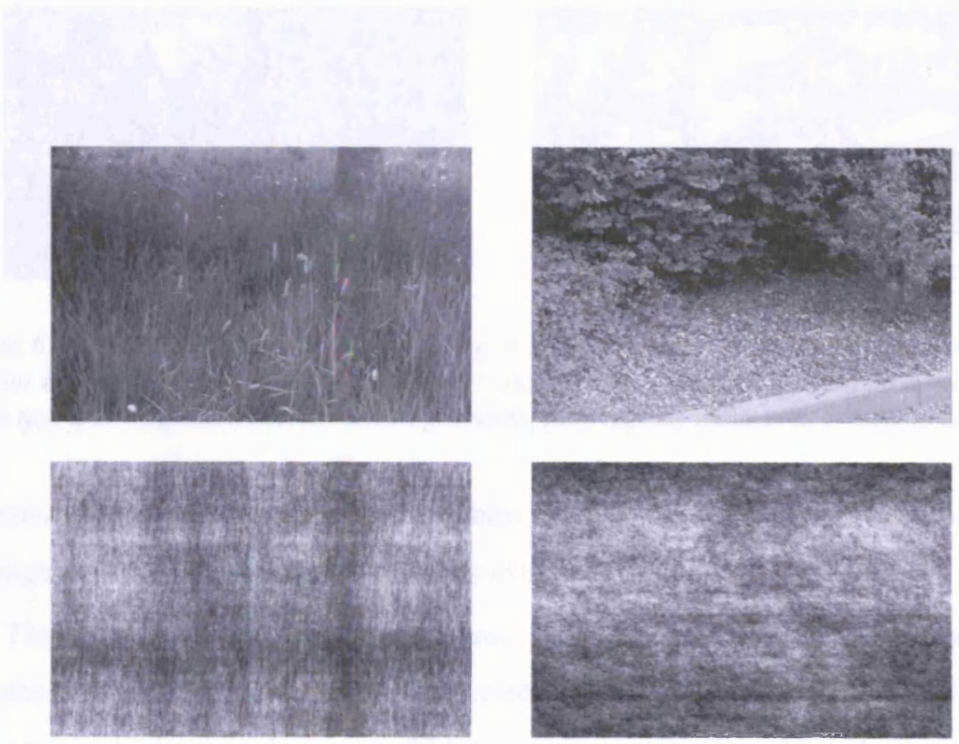
In Figures 6.3(a) and 6.3(b), an individual natural image is shown (Fig. 6.3(a) top left) with its phase randomized version that has an excess of horizontal gradients (Fig. 6.3(b) left). It is also obvious what contents in the natural image, shown in Figure 6.3(a) left, gives rise to the excess of horizontal gradients. Therefore, Figure 6.3 suggests that not every phase randomized natural image has an excess of vertical gradients.

The phase randomized image (Fig. 6.3(a) bottom right) obtained from the natural image in Figure 6.3(a) top right, was deliberately chosen to illustrate contrasting statistics from Figure 6.3(a) bottom left. Figure 6.3(a) bottom right contains an excess of vertical gradients as shown by its histogram of gradient directions (Fig. 6.3(b) bottom right). In this instance, it is less obvious what is causing the excess of vertical gradients. On closer inspection, Figure 6.4 left, which is a region taken from the natural image in Figure 6.3(a) top right, shows how vertical gradients arise from the effect of sunlight—from above—illuminating the leaves of the bush; the gradient directions are encoded as a grey-scale value with white and black representing vertical gradient directions in Figure 6.4 right.

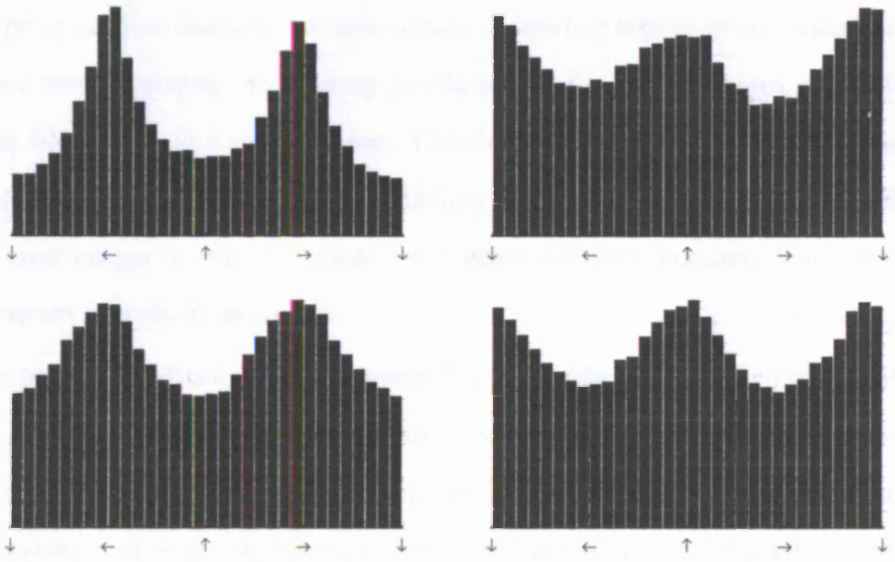
It might be argued that the excess of vertical gradients for an ensemble of phase randomized natural images are artifacts resulting from performing a Fourier transform and inverse Fourier transform, which are both periodic, on images with discontinuities at the top, bottom, left and right sides (i.e. image borders). The boundary effects are caused by artifact edges created by neighbouring pixels on opposite borders of an image.

The difficulty with image border effects is that there does not exist a perfect solution to the problem, although a number of different methods have been adopted in order to deal with them. For example, padding around all four borders of the original image with i) zeros, ii) repeated tiling of the original image, and iii) mirror images of the original. Alternatively, a windowing function can be applied to the image which will soften the strength of any edges created at the image borders.

Nevertheless, in Figures 6.3(a) bottom and 6.3(b) bottom, examples of phase randomized natural images are illustrated that exhibit either an excess of vertical gradients or an excess of



(a)



(b)

Figure 6.3: In (a) left column shows natural image (top) and its phase randomized image (bottom). Illustrated in (b) are the gradient direction histograms (gradient direction indicated by arrow) of the images shown above in (a) arranged in the same order. In the left column of (b) an excess of horizontal gradients are found; in the right column an excess of vertical gradients.

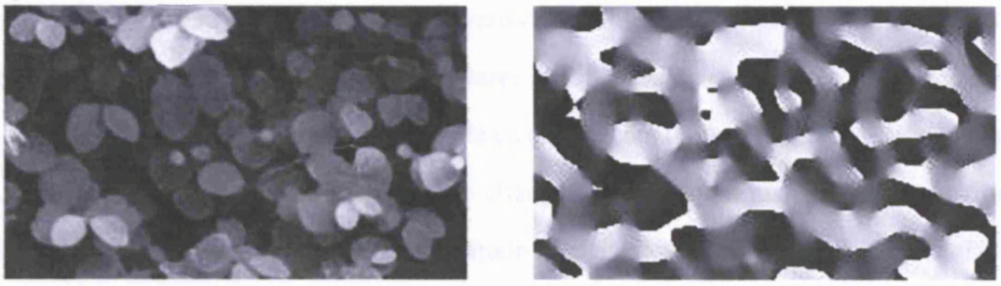


Figure 6.4: Left: illustrates a region from Fig. 6.3(a) top right. Right: illustrates the effect of the sun illuminating the leaves of the bush by encoding the gradient direction on a grey-scale: white and black regions indicate vertical gradients, grey regions indicate non-vertical gradients.

horizontal gradients which have been highlighted previously as being due to image structure in the original natural image, and thus less likely to be artifacts of the phase randomizing process.

The histogram distribution of gradient directions in whitened natural images shows a slight prevalence of vertically and horizontally oriented gradients. This is markedly flatter than the histogram peaks for natural images, which is surprising given that whitened natural images appear to preserve some features of natural images. However, whitened natural images only appear to preserve those features that occur at high contrasting regions in the natural image. But, in whitened natural images, the intensity profile across, for example edges, would be sloped rather than being close to a step function. Therefore, observing weaker peaks in the 1-point gradient direction histograms for whitened images compared to natural images are likely. For Gaussian noise images, as expected, there is no gradient orientation preference (flat distribution) in its histogram of gradient directions.

If we observe the distribution of gradient directions where the gradient magnitude is high (Fig. 6.5) the histogram profiles change for all the image classes except Gaussian noise images. Natural images show a greater prevalence of horizontally oriented gradient directions but no greater prevalence of vertically oriented gradient directions (Fig. 6.5 top left). The increase in horizontal gradients implies the existence of long upright objects in natural images which contrast sharply with that of the background. This may be a consequence of the fact that natural objects, e.g. trees, grow vertically toward the sun.

Phase randomized natural images show increases in the prevalence of both vertically and horizontally oriented gradients (Fig. 6.5 top right) but the increase is most pronounced for vertical. For whitened natural images there is a subtle increase in both the prevalence of horizontally and vertically oriented gradient directions (Fig. 6.5 bottom left). For reasons stated earlier, it is

expected that whitened images converge towards natural images for gradient directions where the gradient magnitude is high as certain features are still visible in whitened natural images.

For Gaussian noise images, no noticeable change occurs in its gradient direction histogram (Fig. 6.5 bottom right); it remains flat. The change in the 1-point histogram distributions of gradient directions where the gradient magnitude is high and when there is no threshold set is different for all four image classes. This suggests that all four image classes behave differently when the gradient magnitude is high. This behaviour shall be observed for 2-point and 3-point dependencies between gradient directions, which are discussed next.

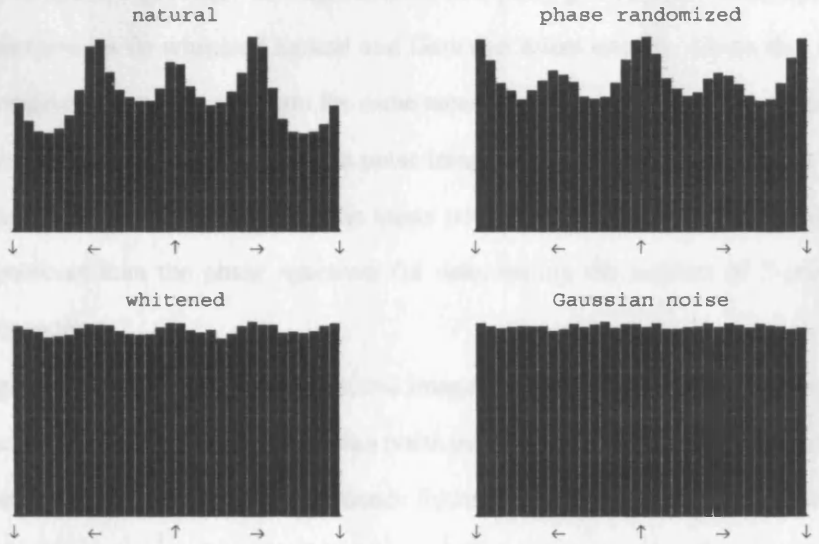


Figure 6.5: Histograms of gradient directions (gradient direction indicated by arrow) where the gradient magnitude is high for the four image classes (natural, phase randomized natural, whitened natural and Gaussian noise images). For natural and phase randomized natural images the distribution has changed significantly from Fig. 6.1: there are more horizontally oriented gradient directions than vertically oriented ones for natural images, and more vertically oriented gradient directions for phase randomized natural images.

6.2 2-point Statistics

In this section the results of computations of 2-point dependencies of gradient directions and magnitudes, and gradient directions where the gradient magnitude is high (strong gradient directions) are presented. Figure 6.6 illustrates the amount of 2-point dependency between certain gradient measurements (magnitude or direction). Along the horizontal axes is the logarithm of the separation between measurements as a multiple k of the first order derivative Gaussian kernel scale $\sigma = 4$ used. The vertical axes is the mutual information shared between the two

gradient measurements in natural units (nats) rather than bits. The error bars are the 95% confidence limits estimated using the bootstrap procedure described in Sections 4.3 and 5.6.

6.2.1 Gradient Direction Dependencies

The top chart in Figure 6.6 shows results for gradient direction dependencies. What is clear is that all image classes show a decrease in mutual information I with increasing separation k between points. This is to be expected; the information shared between points should decrease as the points move further apart. The most significant observation is that natural and phase randomized natural images share the same amount of 2-point gradient direction dependency for a given separation, as do whitened natural and Gaussian noise images. Given that natural and phase randomized natural images share the same mean power spectrum but not phase spectrum, and whitened natural images and Gaussian noise images both have flat mean power spectra but different phase spectra, it suggests that the mean power spectrum of the ensemble of images is more significant than the phase spectrum for determining the amount of 2-point gradient direction dependency.

The agreement in results between natural images and phase randomized images (and between whitened natural images and Gaussian noise images) is to a higher level of accuracy than the error bars which show the 95% confidence limits for the mutual information scores. It is noticeable that the error bars for natural images and phase randomized natural images are larger than either whitened natural images and Gaussian noise images, which are both small. The larger error bars are most likely due to the increased variability in images within that particular image class over another, e.g. natural images compared to Gaussian noise images. It should also be noted that the mutual information, at a given separation, is higher for natural and phase randomized natural images than it is for whitened natural and Gaussian noise images.

6.2.2 Gradient Magnitude Dependencies

For gradient magnitudes, as found for gradient directions, the mutual information for all image classes decreases with the spatial separation between measurements, as shown in the middle chart of Figure 6.6. However, now all four image classes are distinct although overall the dependency between gradient magnitudes is less than it is for gradient directions for a given separation k. For the distance between measurements where $k \leq 1$, natural images have the greatest amount of 2-point mutual information followed by phase randomized natural images,

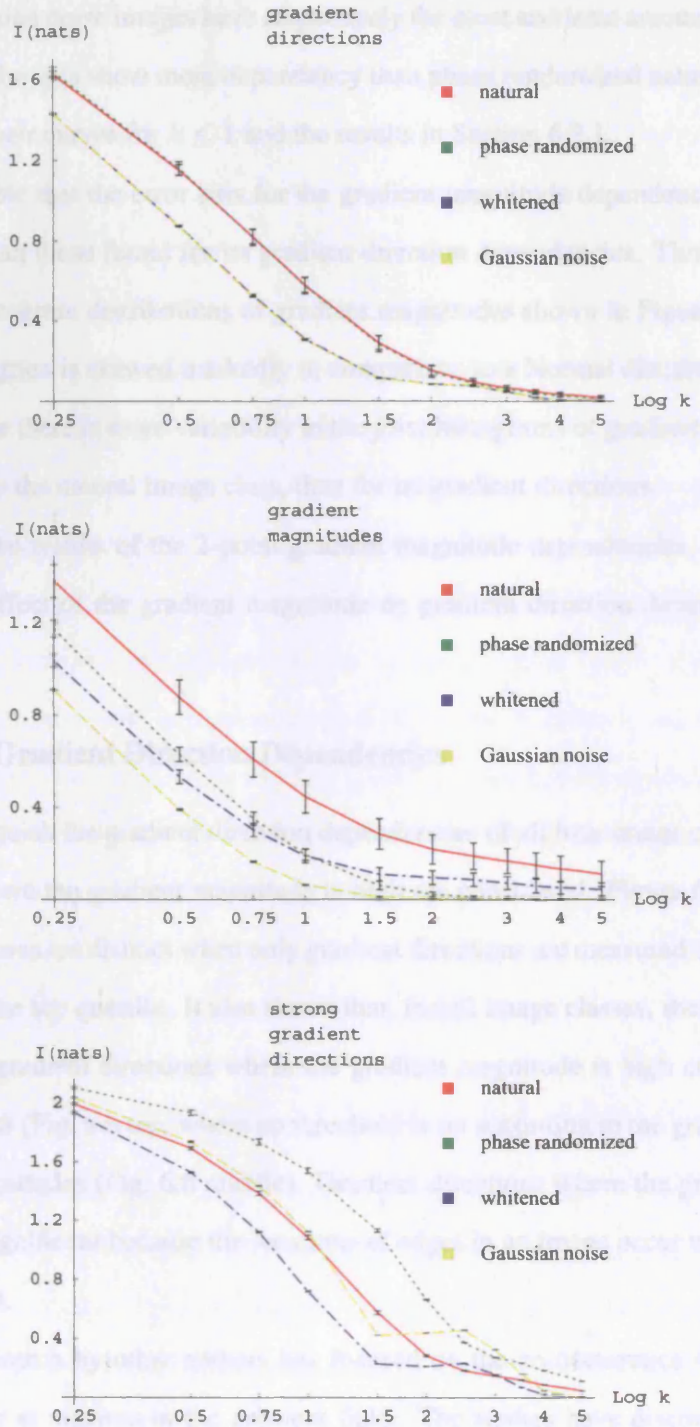


Figure 6.6: Illustrates results of 2-point statistics. Vertical axis: mutual information $I(A;B)$ (nats); horizontal axis: distance (logarithmic scale) between measurements in multiples k of $\sigma = 4$. Key: natural image (red solid line), phase randomized (green dotted), whitened (blue dash-dot), Gaussian noise (yellow dashes). From top to bottom, 2-point mutual information of gradient directions (note that dotted curve is hidden behind solid curve, similarly dash-dot curve hidden behind dashes curve), gradient magnitudes, and gradient directions where the gradient magnitude is high (strong gradient directions).

whitened natural images, and Gaussian noise images. However, for $k > 1$, although natural images and Gaussian noise images have respectively the most and least amounts of dependency, whitened natural images show more dependency than phase randomized natural images, which is in contrast to their curves for $k \leq 1$ and the results in Section 6.2.1.

It is noticeable that the error bars for the gradient magnitude dependencies of natural images are larger than those found for its gradient direction dependencies. This may be a consequence of the histogram distributions of gradient magnitudes shown in Figure 5.7. For natural images, the histogram is skewed markedly in comparison to a Normal distribution. The greater error bars indicate there is more variability in the joint histograms of gradient magnitudes from image to image in the natural image class, than for its gradient directions.

In light of the results of the 2-point gradient magnitude dependencies, it seems sensible to consider the effect of the gradient magnitude on gradient direction dependencies. This is discussed next.

6.2.3 Strong Gradient Direction Dependencies

In order to distinguish the gradient direction dependencies of all four image classes only gradient directions where the gradient magnitude is high are considered. Figure 6.6 bottom, shows that all image classes are distinct when only gradient directions are measured where the gradient magnitude is in the top quartile. It also shows that, for all image classes, the dependencies are greater between gradient directions where the gradient magnitude is high compared to either gradient directions (Fig. 6.6 top, where no threshold is set according to the gradient magnitude) and gradient magnitudes (Fig. 6.6 middle). Gradient directions where the gradient magnitude is high are also significant because the locations of edges in an image occur where the gradient magnitude is high.

Previous research by other authors has focused on the co-occurrence statistics of edges which only occur at maxima in the gradient field. The studies have discovered several regularities between pairs of edges in natural images. For example, it has been shown that the dependency between edges is greatest when the second edge is along the initial edge direction in natural images compared to its (synthetic) phase randomized versions [8]. Another study has found that the most likely contour joining two edge segments that are not iso-oriented is a circular arc and if iso-oriented a line [9]; this result was extended to include that natural scenes are likely to have many closed smooth contours [10]. These previous results suggest that in

natural images the dependency between pairs of edges depends on their relative orientations. However, the results presented here are more general because they are not specific to edges alone. For example, although gradient direction dependencies in natural images increase where the gradient magnitude is high, there is also a greater increase in the dependencies found for phase randomized images. This suggests that it is not only edges which are contributing to the increased dependency between gradient directions.

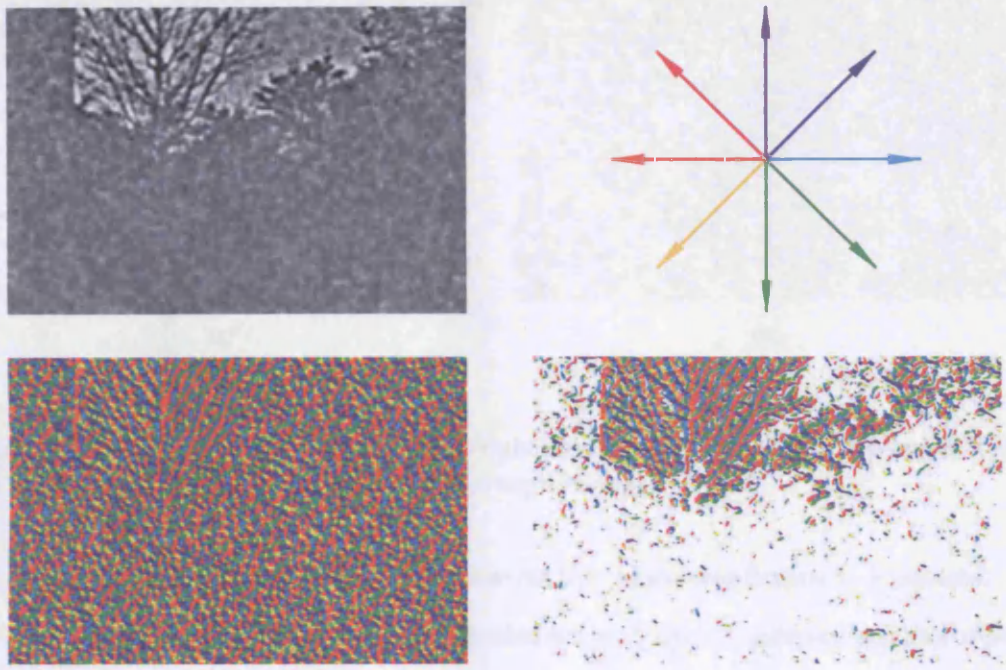
Furthermore, the shape of the curves in Figure 6.6 bottom are different to Figure 6.6 top and middle. For all image classes, the dependency does not fall as sharply for $k \leq 1$. Moreover, for whitened natural images and Gaussian noise images, the dependency increases for $1.5 \leq k \leq 2.5$. This is likely down to bands in the gradient directions. For example, Figures 6.7(a) and 6.7(b), show orientation maps of one whitened natural image and Gaussian noise image, where the gradient direction is colour coded. The banding is most visible in the whitened image.

In order to visualise that a separation of $k = 2.5$ is similar to the width of the banding see Figures 6.8(a) and 6.8(b) which show black circles of diameter equal to a separation of $k = 2.5$. In fact, the effect of this banding also appears in the dependency between the gradient directions in Figure 6.6. For $1.5 \leq k \leq 2.5$ there is a very subtle levelling off in the curves for both whitened natural images and Gaussian noise images. This suggests that gradient directions where the gradient magnitude is high emphasises the banding effect for $1.5 \leq k \leq 2.5$, resulting in an increase in the dependency between gradient directions.

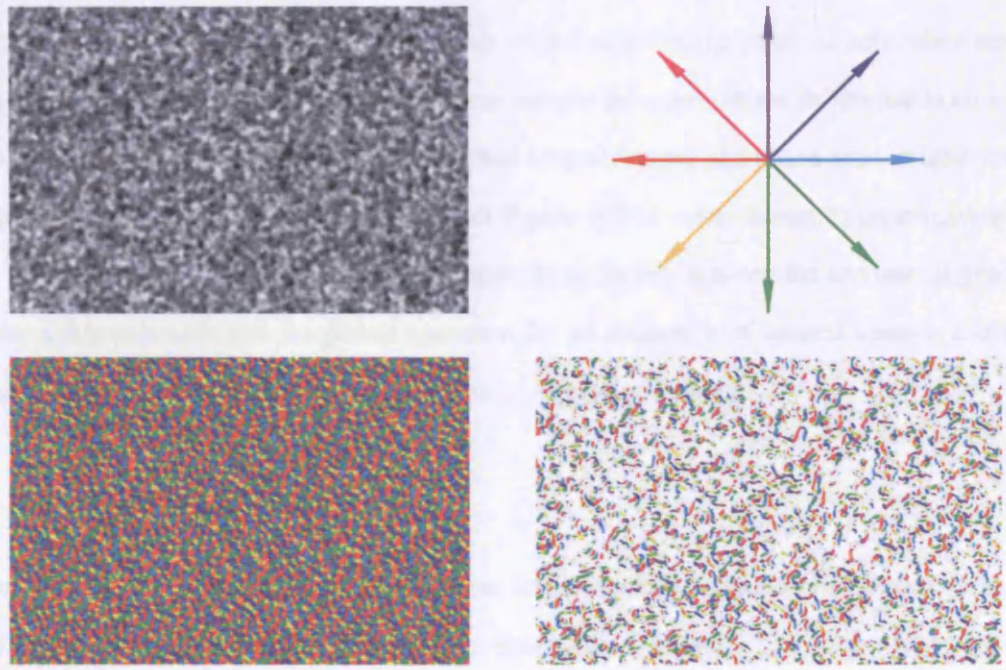
6.3 Gradient Direction Dependencies of Individual Natural Images

So far, the results presented have been mainly for ensembles of images. In this section, gradient direction dependencies of individual natural images are compared with that of their phase randomized versions (same power spectra as original natural image but with random phases). Two examples of an individual natural image and its phase randomized version are shown in Figure 6.9(a).

From the two natural images in Figure 6.9(a) the sampling of gradient directions—as described in Section 5.3 but with 100,000 gradient directions per image rather than 10,000—was repeated ten times for different separations k. The 2-point and 3-point gradient direction dependencies are then calculated for each sample, at each k, for each image. For phase randomized

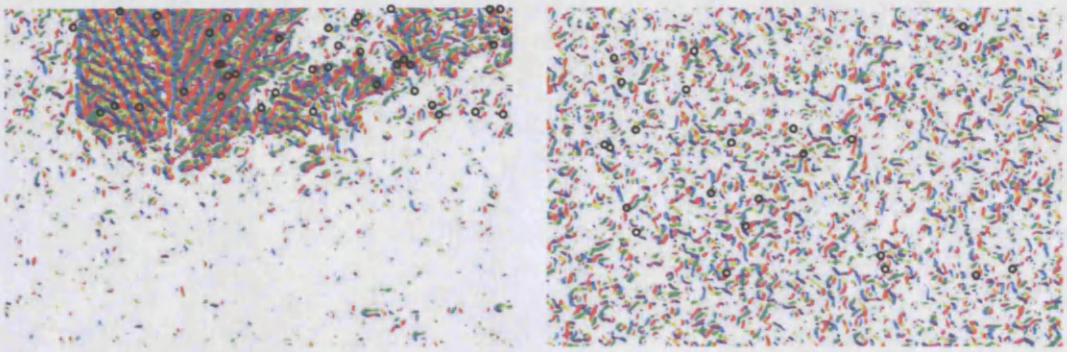


(a)



(b)

Figure 6.7: For both whitened image in (a) and Gaussian noise image in (b), the gradient direction has been encoded as a colour (see key top right). Bottom left of (a) and (b): the image top left has been converted to a colour map of gradient directions. Bottom right of (a) and (b): same as bottom left, but with only the gradient directions where the gradient magnitude is high.



(a) (b)

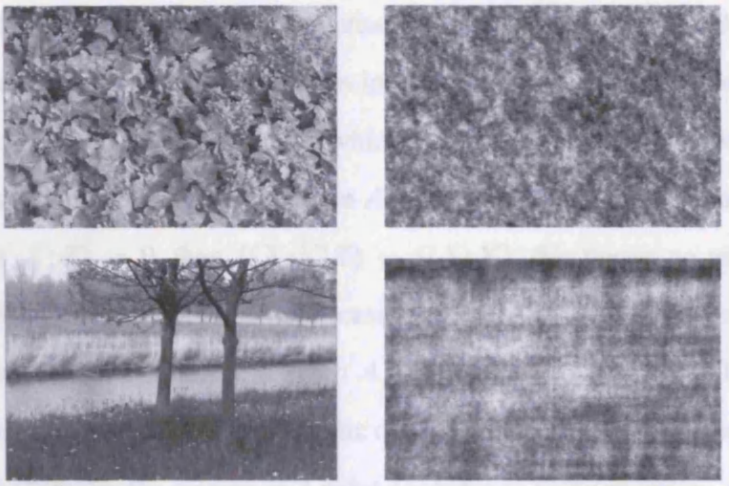
Figure 6.8: Figures (a) and (b) are the bottom right illustrations in Figures 6.7(a) and 6.7(b) but with black circles of diameter equal to $k = 2.5$ superimposed.

images, the same is done, but each time a new random phase distribution is generated. Figure 6.9(b) shows, at each k, the mean information score of the ten samples together with the 95% confidence limits estimated from the standard error in the mean.

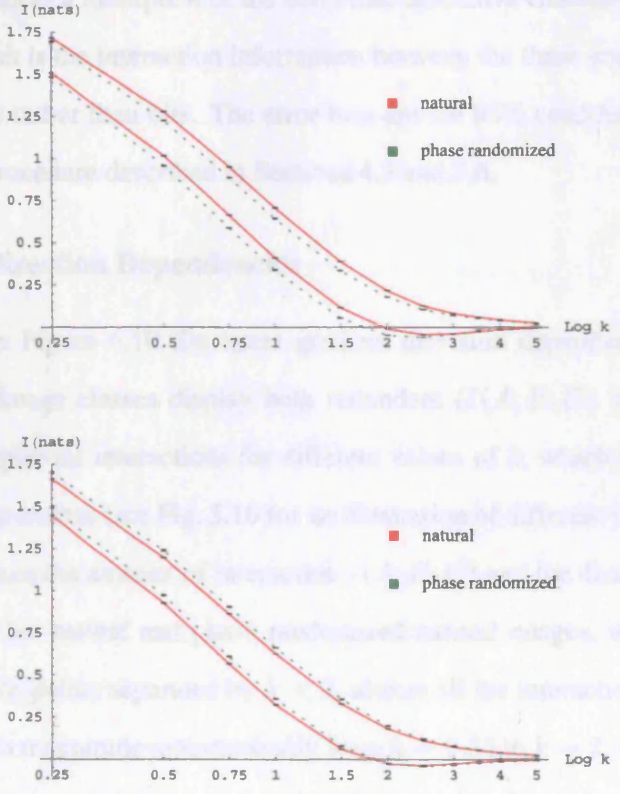
The results show that there are differences, which are significant but small, in the amount of dependencies of gradient directions between an individual image and the equivalent natural image with randomized phases. Further, it is not simply the case that the difference is either always smaller or always larger between individual natural images and phase randomized natural images; the two example images illustrated in Figure 6.9(b) were chosen to demonstrate that both possibilities do occur. Therefore, the relationship found between the amount of gradient direction dependencies and the power spectrum for an ensemble of natural images and their phase randomized versions cannot be extended to individual images.

6.4 3-point Statistics

In this section the results of computing 3-point dependencies of gradient magnitudes and gradient directions are presented. The focus in this section is on the triplewise interaction of three gradient measurements A–B–C. Recall that the interaction information of three gradients $I(A; B; C)$ describes how the mutual information between A and B reduces in the context of C. $I(A; B; C)$ describes information common to A, B, and C but which is not present in any subset, e.g. A–B, B–C, A–C. In terms of probability theory, even if $p(A, B, C) = p(A, B)p(C)$, this will result in no contribution to $I(A; B; C)$ because only A and B are dependent but A, B, and C together are independent.



(a)



(b)

Figure 6.9: In (a), left column shows natural images and right column their respective phase randomized image. In (b) results from natural image (red solid) and phase randomized (green dotted); the error bars on each curve indicate the 95% confidence limits. The upper pair of curves in (b) represent 2-point gradient direction mutual information score $I(A;B)$, and the lower pair the 3-point $I(A;B;C)$. The top graph show results from the top images in (a), and the bottom graph from bottom images.

Recall that positive values of 3-point interaction information $I(A;B;C)$ indicate redundancy. If $I(A;B;C) > 0$ then $I(A;B)$ (mutual information between A and B), which describes how the uncertainty in A reduces when informed by B, must also be positive. Furthermore, if $I(A;B;C) > 0$, then $I(A;B|C)$, which is the conditional mutual information that describes how the uncertainty reduces between A and B in the context of C, must be less than $I(A;B)$. If $I(X;Y;Z) = 0$, then $I(X;Y|Z) = I(X;Y)$, this means no triplewise dependencies exist. Moreover, $I(A;B;C)$ is not necessarily nonnegative. This suggests that there is a further reduction in the uncertainty between A and B in the context of C. If $I(A;B;C)$ is negative then this is said to indicate a synergetic dependency between A, B, and C.

Figure 6.10 illustrates the amount of 3-point dependency between triples of gradient measurements (magnitude or direction). Along the horizontal axes is the logarithm of the separation between measurements as a multiple k of the first order derivative Gaussian kernel scale $\sigma = 4$ used. The vertical axes is the interaction information between the three gradient measurements in natural units (nats) rather than bits. The error bars are the 95% confidence limits estimated using the bootstrap procedure described in Sections 4.3 and 5.6.

6.4.1 Gradient Direction Dependencies

The top left graph in Figure 6.10 illustrates gradient direction dependencies. The first observation is that all image classes display both redundant ($I(A;B;C) > 0$) and synergetic ($I(A;B;C) < 0$) triplewise interactions for different values of k, which is the separation between gradient measurements (see Fig. 5.10 for an illustration of different values of k). Second, the relationship between the amount of interaction $I(A;B;C)$ and the distance between points k is almost identical for natural and phase randomized natural images, which is the same as the 2-point results. For points separated by $k < 2$, almost all the interaction is redundant. The redundancy reduces in magnitude monotonically from $k = 0.25$ to $k = 2$. Crossover from positive to negative values occurs between $1.5 < k < 2$, and we find less than 0.1 nats of synergy for separations between $1.5 < k < 4$.

Whitened natural and Gaussian noise images share an almost identical relationship between the amount of dependency between gradient directions and the distance between points. For points separated by $k < 1$, almost all the interaction is redundant. The redundancy reduces in magnitude monotonically from $k = 0.25$ to $k = 1$. Crossover from positive to negative values occurs between $1 < k < 1.5$ and less than 0.1 nats of synergy are found for $1 < k < 2.5$.

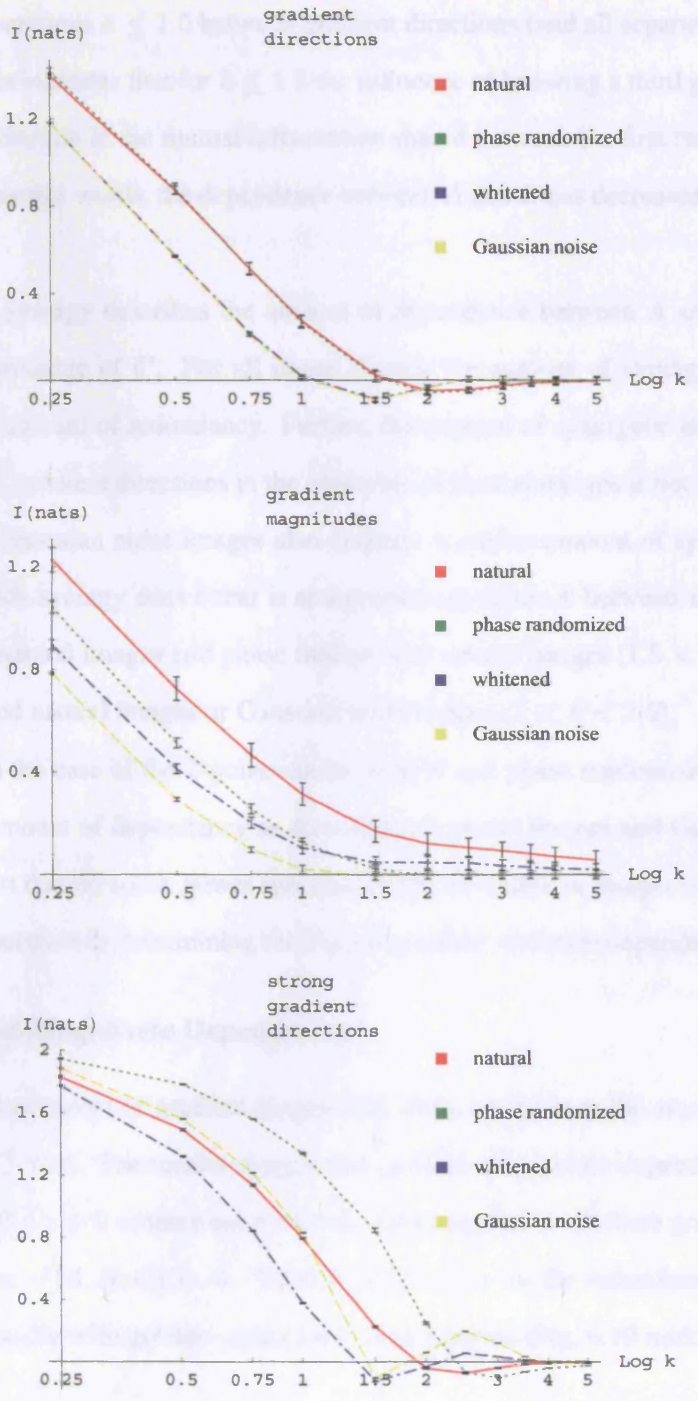


Figure 6.10: Illustrates results of 3-point statistics for gradient direction (top), gradient magnitude (middle), and gradient directions where the gradient magnitude is high (bottom) in different image classes. Vertical axis: interaction information $I(A;B;C)$ (nats), Horizontal axis: distance (logarithmic scale) between measurements in multiples k of σ. Key: natural image (solid line), phase randomized (dotted), whitened (dash-dot), Gaussian noise (dashes), note that dotted curve is hidden behind solid curve, similarly dash-dot curve hidden behind dashes curve. If $I(A;B;C) < 0$ (> 0) there is said to be synergy (redundancy).

In summary, the natural image class shows redundancy (positive values of information in Fig. 6.10) for separations $k \leq 1.5$ between gradient directions (and all separations for gradient magnitudes). This indicates that for $k \leq 1.5$ the influence of knowing a third gradient direction C, results in a reduction in the mutual information shared between the first two gradient directions A and B. In other words, the dependence between A and B has decreased with knowledge of C.

In contrast, synergy describes the amount of dependence between A and B that has increased with knowledge of C. For all image classes the amount of synergy found is small compared to the amount of redundancy. Further, the amount of synergetic dependency found between triples of gradient directions in the ensemble of natural images is not unique given that the ensemble of Gaussian noise images also displays a similar amount of synergy. However, the region in which synergy does occur is at a greater separation k between measurements for the ensemble of natural images and phase randomized natural images ($1.5 < k < 4$) than it is for either whitened natural images or Gaussian noise images ($1 < k < 2.5$).

Lastly, as in the case of the 2-point results, natural and phase randomized natural images share the same amount of dependency as do whitened natural images and Gaussian noise images. This suggest that the mean power spectrum of the ensemble of images is more significant than the phase spectrum in determining the 3-point gradient direction dependencies of images.

6.4.2 Gradient Magnitude Dependencies

Notably, the dependencies of gradient magnitudes show no synergy for any image class, i.e. where $I(A;B;C) < 0$. The results suggest that gradient magnitude dependencies are either redundant $I(A;B;C) > 0$ or there are no interactions common to all three gradient magnitude measurements, i.e. $I(A;B;C) = 0$. When $I(A;B;C) > 0$, the redundant information decreases monotonically with greater separation between points (Fig. 6.10 middle) for all image classes.

Furthermore, as in the case of 2-point dependencies, all image classes are distinct for gradient magnitude dependencies. For $k \leq 1$, natural images have the greatest amount of redundant 3-point interaction information followed by phase randomized natural images, whitened natural images and Gaussian noise images. For $k > 1$, natural images show the most redundant interaction information but this is now followed by whitened natural images, phase randomized natural images and lastly Gaussian noise images. This ordering follows results found for the

computation of 2-point dependencies in Section 6.2.2.

6.4.3 Strong Gradient Direction Dependencies

Again it is possible to distinguish the gradient direction dependencies of all four image classes by only computing the dependencies between gradient directions where the gradient magnitude is in the top quartile. This results in a significant change in the way the amount of redundancy between gradient directions falls with separation k. For gradient directions where the gradient magnitude is high, the amount of redundant interaction information is higher for a given k and the point at which the curves for all four image classes reaches zero is for greater k compared to the dependency of gradient directions where no threshold is set according to the gradient magnitude.

Significantly, for natural images, the region in which synergy occurs is $2 \leq k \leq 4$. This is within the same region as found for gradient directions (Fig. 6.10 top) $1.5 \leq k \leq 4$. However, for phase randomized natural images, the region of synergy is now $2.5 \leq k \leq 5$, which is a shift to greater separations compared to previous results.

6.5 Discussion of Main Results: 1-, 2-, 3-point Statistics

The 1-point results consist of histograms of gradient directions found in the four image classes: natural, phase randomized natural, whitened natural and Gaussian noise. It has been found that natural images contain a prevalence of vertical and horizontal gradients, and furthermore, there are more upward vertical gradients than downward. It has been argued that this bias is due to natural images being composed of sky (lighter) and ground (darker) regions. This contrast between the upper and lower regions results in vertical upward gradients. For phase randomized natural images, there is a prevalence of vertical gradients which has also been attributed to the sky and ground regions and that lighting in natural scenes, on average, is from above due to sunlight and skylight. Whitened natural images have a weaker prevalence of vertical and horizontal gradients compared to natural images. Gaussian noise images, as expected, have a flat distribution of gradient directions.

If only gradient directions where the gradient magnitude is high are examined, then the histogram for natural images changes to having more horizontal gradients than vertical. Phase randomized natural images have an even greater prevalence of vertical gradients and an increase in horizontal gradients. Whitened natural images show a slight increase in horizontal

and vertical gradients, and Gaussian noise images remain flat.

The 2-point and 3-point results show that the amount of gradient direction dependency and its variation with separation between points k for an ensemble of natural images is the same as for an ensemble of phase randomized natural images. It is also found that whitened natural images (created from a flat power spectra, and phase spectra from natural images) have the same amount of 2-point and 3-point gradient direction dependency as Gaussian noise images although both have less dependency than is found for natural images. These results suggest that for an ensemble of natural images the average dependencies between 2-point and 3-point gradient directions are dependent upon the ensemble's mean power spectrum rather than the phase spectra of natural images.

If instead either the dependencies of i) gradient directions where the gradient magnitude is high, or ii) gradient magnitudes are computed, then all four image ensembles (natural, phase randomized natural, whitened natural and Gaussian noise images) have distinct results. This observation suggests that there are, at least, some differences between the gradient direction dependencies in natural images and phase randomized images because the contributions from gradient directions where the gradient magnitude is high are different. Thus, it can be said that the relationship found between the amount of gradient direction dependencies and the mean power spectrum of an ensemble of natural images cannot be extended to the dependencies between gradient directions where the gradient magnitude is high, or between gradient magnitudes. It has also been shown that the amount of gradient direction dependency for an individual natural image is different to its phase randomized natural image. Therefore the result does not generalise to individual natural images but remains a property of only an ensemble of natural images.

In summary, the main results presented suggest that for an ensemble of natural images the amount of dependency between 2-point and 3-point gradient directions is determined by the ensemble's mean power spectrum. This result is not trivial, i.e. it is not a more general property of first-order derivative measurements, given that the 2-point and 3-point gradient magnitude dependencies from an ensemble of natural images is distinguishable from its phase randomized ensemble as are dependencies between gradient directions where the gradient magnitude is high. Furthermore, the relationship does not hold between individual natural images and phase randomized natural images.

The main results also imply that the phase spectrum of an ensemble of natural images, on average, does not contribute to the amount of gradient direction dependency anymore than if the natural images contained random phases. This is surprising because the phase information of natural images has been shown to contribute to localised features such as lines and edges in natural images (Sec. 3.5.2).

Alterations to the phase information degrades a natural image giving an 'unnatural' appearance. Therefore, it was expected that the amount of gradient direction dependencies of an ensemble of phase randomized images would at least be different to natural images. However, in Section 3.3.4, it was described how the power spectrum can, for certain images which are rich in textures and shading or with strong geometric forms, contribute significantly to the appearance of an image. Therefore, given that gradient direction dependencies have been examined rather than the orientation of edges, it is not necessary for an ensemble of images to retain line and edge structure in order to keep, on average, the same amounts of gradient direction dependency.

An example of a natural image with a distinct shading pattern and its phase randomized counterpart is presented in Figure 6.11. The phase randomized natural image (Fig. 6.11 right) looks like the shading pattern of Figure 6.11 left.

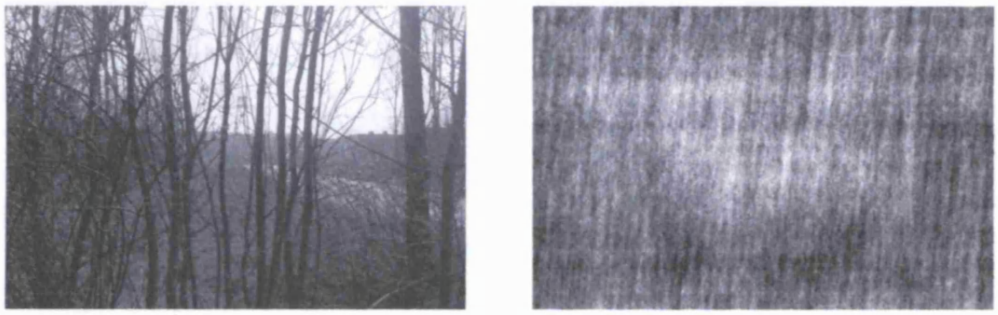


Figure 6.11: Left: natural image with an interesting shading pattern. Right: phase randomized version of natural image (left) showing that the shading pattern is retained but the localisation of any lines and edges is lost.

Next, in Chapter 7, further experiments are presented which aim to explore in more detail the results found in this chapter. The additional experiments conducted include examining the gradient direction dependencies of images with varying power law power spectra and random phases, and images generated from the dead leaves model.

Furthermore, the 2-point gradient direction dependencies are reported with respect to not

only distance but the relative angle between gradient directions. Gradient direction dependencies are also computed for an in-a-line spatial configuration of points rather than a triangular configuration. Finally, an explanation for the presence of synergetic dependencies between 3-point gradient directions is presented.

Chapter 7

Further Experiments

In order to clarify and generalise the main results presented in Chapter 6, additional experiments are presented here.

In Section 7.1.1, investigations are carried out to see whether there exists a lawful dependence between the decay rate of the gradient direction dependencies and α the decay constant of the ensemble power spectrum which is of the form $1/f^{\alpha}$. In Section 7.1.2, it is demonstrated that an ensemble of images generated from the dead leaves model has the same amount of gradient direction dependencies as an ensemble of natural images. In Section 7.1.3, an example of an image class (constant-sized disks) which has different gradient direction dependencies compared to its phase randomized version is presented. This image class is a modified version of the dead leaves model which has scale-variant gradient direction dependencies.

In Section 7.2 not only is the separation between two gradient directions considered but also the effect of the relative angle between the gradient directions.

In Sections 7.3.1 and 7.3.3, the synergetic gradient direction dependencies are explored in more detail. For instance, the effects of choosing a triangular spatial configuration of gradient measurements is considered by comparing results computed from a line spatial configuration of measurements. Further, the hypothesis that synergetic gradient direction dependencies arise in images because the curl of any scalar field has zero curl is also investigated.

7.1 Analysis of Other Synthetic Images

In the following section the results from the analysis of other classes of synthetic images are presented together with the methodology to generate each image class.

7.1.1 Varying Power Spectral Slope Images

In the main results (Sections 6.2, and 6.4) it was observed that ensembles of natural images and phase randomized natural images contain the same amount of 2-point and 3-point gradient direction dependencies. Thus, it seems appropriate, because both those image classes have an approximate $1/f^2$ mean ensemble power spectra, to study other classes of images with varying power spectral slopes of the form $S(f) \propto 1/f^{\alpha}$. The significance of images with this power law power spectra was discussed in Section 3.3.1.

Images were generated with varying power law dependency between the spatial frequency f, and their power spectra $S(f)$. In the spatial frequency domain, the phases are set to random values on the interval $[-\pi, \pi]$ whilst respecting conjugate symmetry, and the power spectra are set to follow a power law of the form: $S(f) \propto 1/f^{\alpha}$ where

$$\alpha \approx 3, 2, 1.5, 1, 0, -1, -2, -3$$

and then the images are inverse Fourier transformed to the spatial domain. Recall that natural images have $\alpha \approx 2$ and Gaussian noise images have no spatial frequency dependency (flat) power spectra, $\alpha \approx 0$. Examples of noise images with $\alpha \approx 3, 1, -1, -3$ are illustrated in Figure 7.1. For greater values of α the image appears 'smoother'; this is to be expected because less high frequency energy is present.

The amount of 2-point and 3-point dependency between gradient directions in varying power law power spectra images are shown in Figures 7.2 and 7.3 respectively. It shows that for all classes of image, an increase in separation between the gradient direction measurements results in a decrease in the amount of dependency between the gradient directions.

Figures 7.2 and 7.3 also show that a positive increase in α ($S(f) \propto f^{\alpha}$) decreases the 2-point $I(X;Y)$ and 3-point $I(X,Y,Z)$ mutual information between gradient directions.

For small separations k between the locations where gradient directions are extracted, we can quantify the relationship between α and the decay rate of gradient direction dependencies. The curves in Figure 7.2 produce straight lines when plotted with the vertical and horizontal axes scaled to natural logarithm and linear scales respectively, as shown in Figures 7.4 and 7.5.

Figures 7.4 and 7.5 show that for small separations k between gradient directions both the total 2-point and 3-point gradient direction dependencies decay exponentially with greater

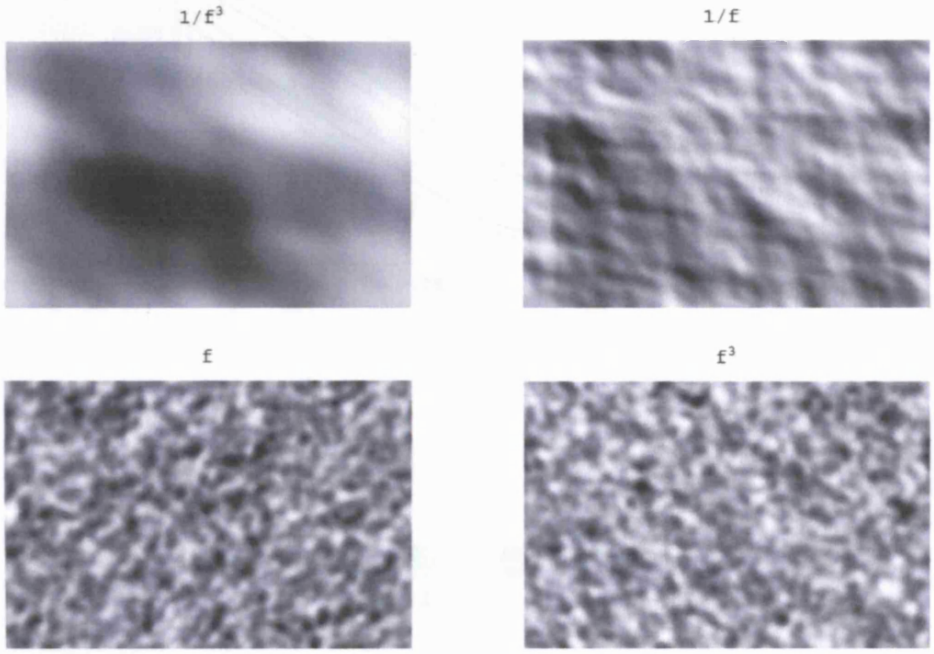


Figure 7.1: Illustrates examples of images with different power law scaling of their power spectra with respect to spatial frequency. All images are blurred by convolution with a zeroth-order Gaussian of scale $\sigma = 4$. All images have random phase spectra but different power spectra. Top row shows images that have power spectra with $1/f^3$ and $1/f$ dependence respectively; bottom row shows images that have power spectra with f and f^3 dependence respectively.

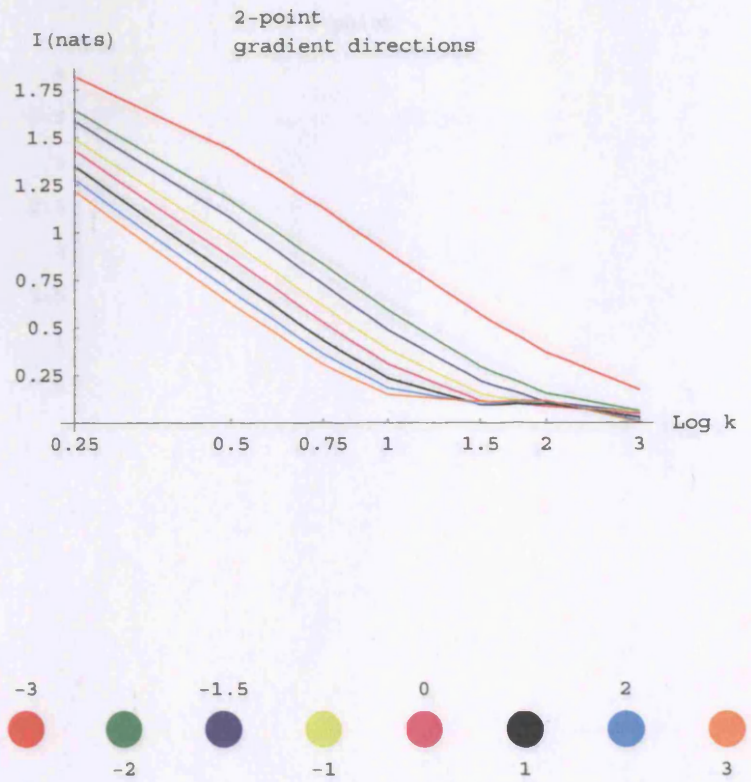


Figure 7.2: Illustrates total 2-point gradient direction dependencies $I(A;B)$ of images with varying power law power spectra and random phases. Vertical axis: mutual information I(nats); horizontal axis: distance (logarithmic scale) between measurements in multiples k of σ. Each curve represents the amount of 2-point dependency $I(A;B)$ between gradient directions of noise images with varying power law power spectra $S(f) \propto f^{\alpha}$. From the top to bottom curve $\alpha \approx -3, -2, -1.5, -1, 0, 1, 2, 3$.

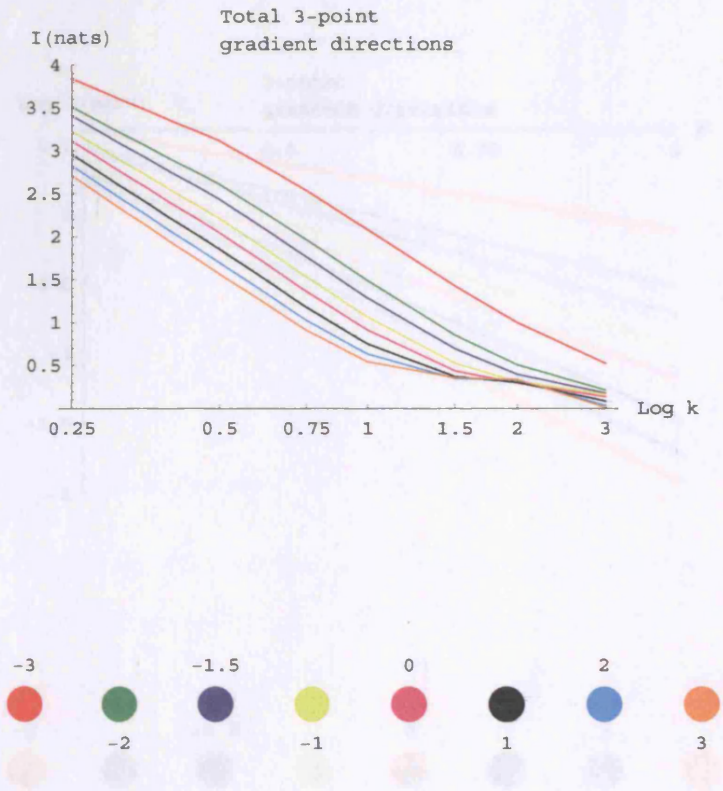


Figure 7.3: Illustrates total 3-point gradient direction dependencies $I(A, B, C)$ of images with varying power law power spectra and random phases. Vertical axis describes the amount of 3-point gradient direction dependencies. Horizontal axis: distance (logarithmic scale) between measurements in multiples k of σ. Each curve represents the amount of 2-point dependency $I(A; B, C)$ between gradient directions of noise images with varying power law power spectra $S(f) \propto f^{\alpha}$. From the top to bottom curve $\alpha \approx -3, -2, -1.5, -1, 0, 1, 2, 3$.

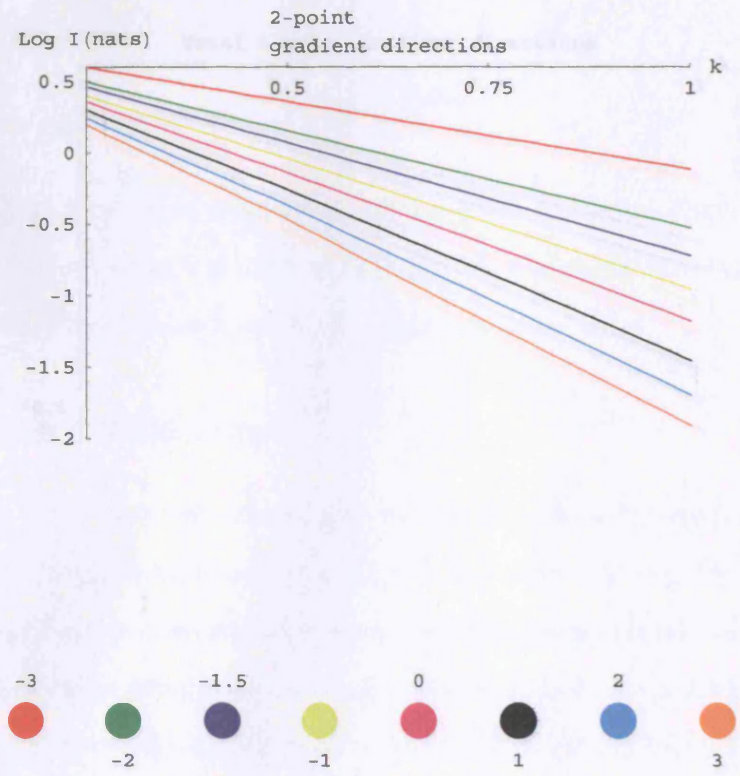


Figure 7.4: Illustrates the total 2-point gradient direction dependencies $I(A;B)$ of curves in Fig. 7.2 on a natural logarithm versus linear axes.

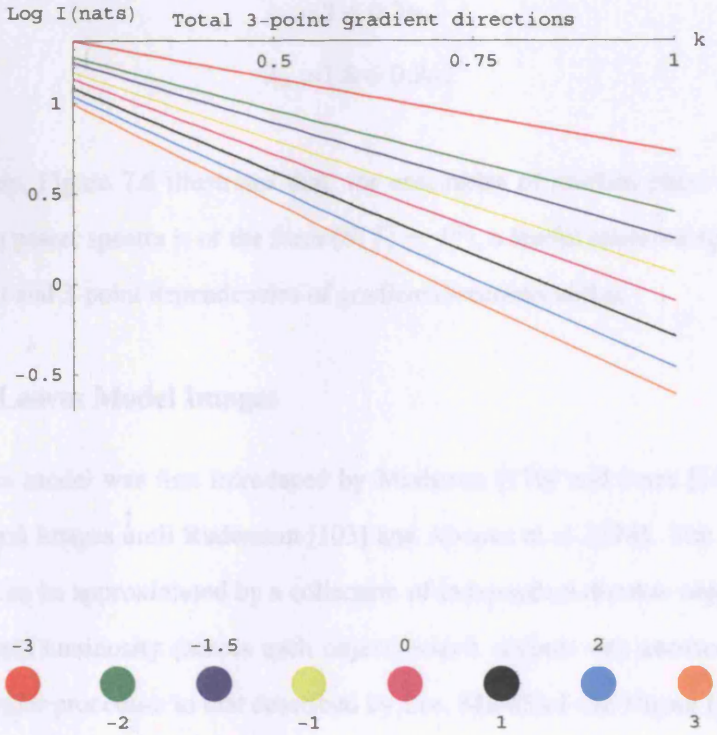


Figure 7.5: Illustrates the total 3-point gradient direction dependencies $I(A, B, C)$ of curves in Fig. 7.3 on a natural logarithm versus linear axes.

distance between measurements. Thus, for $k \leq 1$, we have:

$$I(X;Y) \propto \exp^{-\lambda_1 k} \tag{7.1}$$

$$I(X,Y,Z) \propto \exp^{-\lambda_2 k} \tag{7.2}$$

and approximately, we find that λ_1 and λ_2 satisfy:

$$\lambda_1 = 2 + 0.3\alpha \tag{7.3}$$

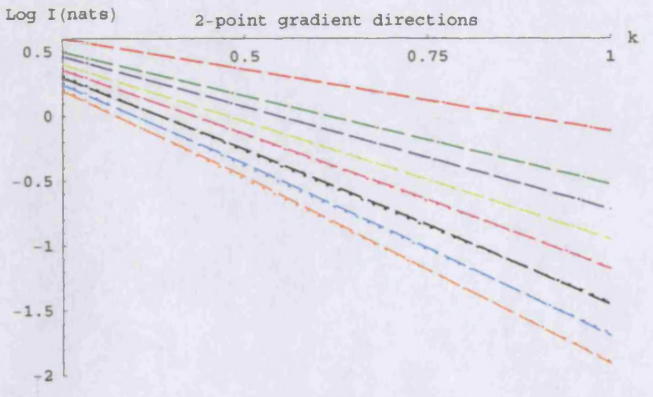
$$\lambda_2 = 1.5 + 0.2\alpha. \tag{7.4}$$

In summary, Figure 7.6 illustrates that, for ensembles of random phase spectra images where the mean power spectra is of the form ($S(f) \propto f^{\alpha}$), a lawful relationship exists between the total 2-point and 3-point dependencies of gradient directions and α.

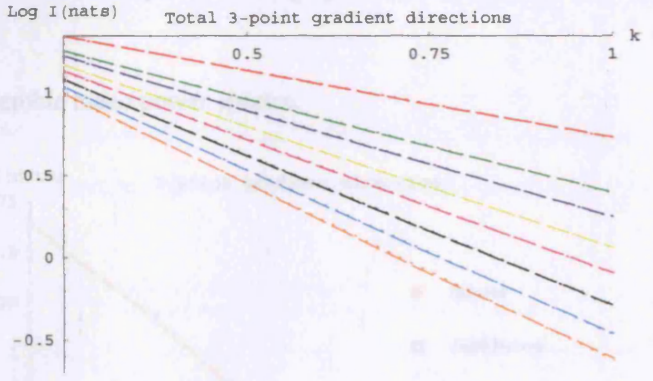
7.1.2 Dead Leaves Model Images

The dead leaves model was first introduced by Matheron [176] and Serra [177] but was not applied to natural images until Ruderman [103] and Alvarez et al. [178]. The model assumes that the world can be approximated by a collection of independent discrete objects of different sizes but constant luminosity (across each object) which occlude one another. In this study we follow a similar procedure to that described by Lee, Mumford and Huang [179] to produce images which model the scale-invariant property of an ensemble of natural images and therefore its $1/f^2$ ensemble mean power spectrum.

Specifically, a $1/r^3$ distribution of disk sizes is used with $1/8 \leq r \leq 2048$. The luminance value of each disk is drawn randomly from a histogram of luminance values taken from an ensemble of natural images (the same ensemble of natural images taken from the van Hateren database is used). The disks are progressively placed on to a plane of dimensions 4096 by 4096 until the whole area is covered. The images are then cropped to match the dimensions of the ensemble of natural images which are 1024 by 1536. An example of an image generated from the dead leaves model is shown in Figure 7.7. Figure 7.8 shows that the amount of 2-point gradient direction dependency for the ensemble of natural images is approximately the same as the dead leaves model images. This result supports the hypothesis that the amount of gradient direction dependencies will be the same for two ensembles of images which have approximately



(a)



(b)

(c)

Figure 7.6: Illustrates the relationship between α (exponent of the power law scaling of the power spectrum) and λ_1, λ_2 the decaying exponents in Eqns. 7.1 and 7.2. Dashed curves in (a) and (b), from top to bottom, represent the curves in Figures 7.4 and 7.5 for images with power spectra ($S(f) \propto f^{\alpha}$) $\alpha \approx -3, -2, -1.5, -1, 0, 1, 2, 3$, see (c) for colour key. Dotted curves next to each dashed curve are plots of Eqns. 7.1 and 7.2 with values given by given by Eqns. 7.3 and 7.4.

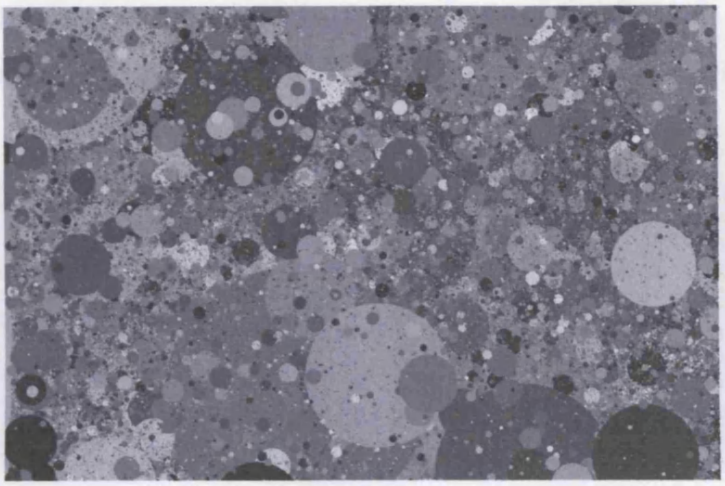


Figure 7.7: An example of an image generated by the dead leaves model.

the same $1/f^2$ ensemble mean power spectra.

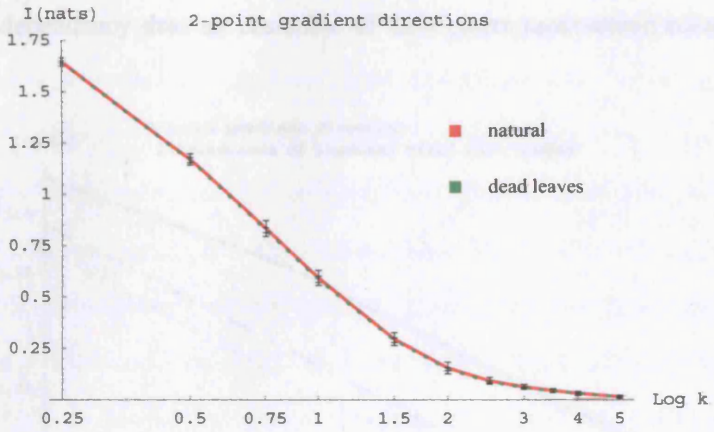


Figure 7.8: Shows 2-point dependencies between gradient directions for natural images (solid line) and dead leaves model images (dotted line which lies just underneath solid line).

7.1.3 Constant-Sized Disk Images

The dead leaves model, described in the previous section (Section 7.1.2), is now modified by using disks with constant sizes of radius $r = 40$, illustrated in Figure 7.9, rather than a $1/r^3$ distribution. Random values are used for the luminance value of each disk. The mean power spectrum of an ensemble of these images does not follow a power law and is therefore not scale-invariant. Scale variance was confirmed by repeating our computations of gradient direction dependencies at twice the previous scale, 2σ. Recall that the image distance between

measurements is also scaled by σ, therefore, for a given separation k between gradient directions, the amount of gradient direction dependency should be constant if scale invariance is to be satisfied.

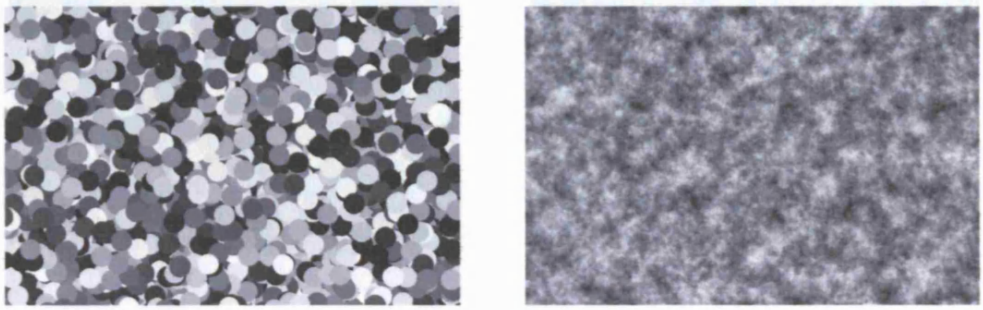


Figure 7.9: Left, an example of a constant-sized disk image; right, a phase randomized constant-sized disk image.

Figure 7.10 shows that an ensemble of constant-sized disk images has more 2-point gradient direction dependency than an ensemble of their phase randomized counterpart images.

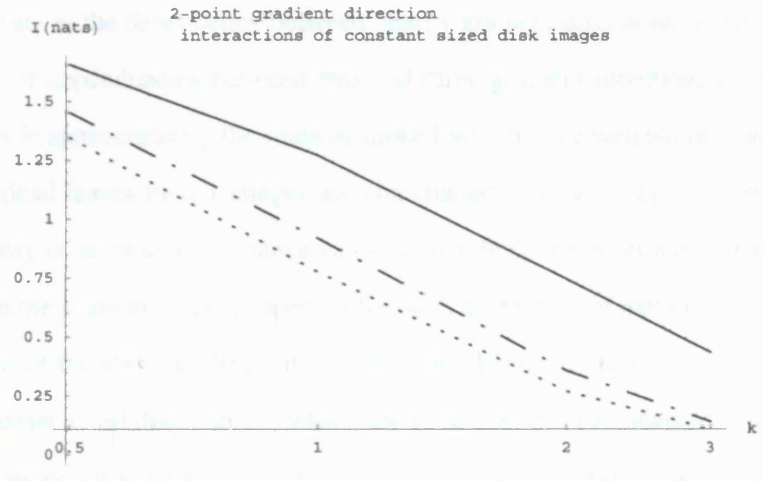


Figure 7.10: Illustrates comparison of 2-point gradient directions interactions for constant-sized disk images (solid), their phase randomized counterparts (dotted), and constant-sized disk images at twice the scale (dash-dot).

If we examine the gradient directions of Figure 7.11 as a colour encoded direction map, then the phase randomized version does appear less ordered than the original constant-sized disk image and therefore this result is not surprising.

Figure 7.10 also shows that the amount of 2-point gradient direction dependencies change depending on the scale of measurement. The amount of gradient direction dependency at scale

σ is different to that at scale 2σ. The pattern of these results is also repeated for the 3-point gradient direction dependencies (results not shown).

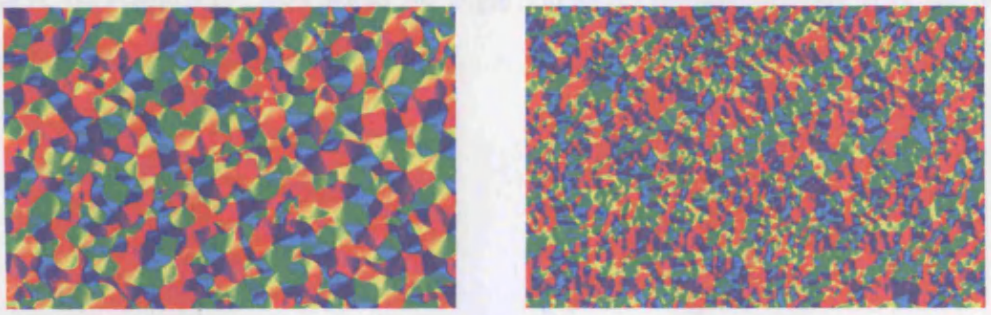


Figure 7.11: Illustrates the maps of gradient directions encoded as a colour for images in Figure 7.9. Left, are the gradient directions for the constant-sized disk image; right, the phase randomized constant-sized disk image.

7.1.4 Conclusion of Further Experiments on Synthetic Images

Ensembles of images with varying power law power spectra, show an approximate lawful relationship between the power law exponent of the ensemble's mean power spectrum and the spatial decay rate in the dependency between nearby gradient directions. Furthermore, the average amount of dependencies between two and three gradient directions of an ensemble of natural images is approximately the same as those found in an ensemble of dead leaves model images. The dead leaves model images are constructed from a collage of disks with a cubic power law decay of sizes and luminance values drawn from the ensemble of natural images in order to match the scale-invariant properties found in ensembles of natural images.

In support of the above findings, it has also been demonstrated experimentally that an ensemble of constant-sized disk images, which are generated from an adapted dead leaves model that does not maintain scale invariance and a power-law form power spectrum, shows more gradient direction dependencies compared to its phase randomized version.

7.2 Information Contour Map of 2-point Statistics

Thus far, the dependencies between 2-point and 3-point gradient directions have been described as a function of distance between measurements only, however, it is also possible to compute dependencies as a function of angle between measurements. For example, in the 2-point configuration, it is possible to compute how much information is shared between a gradient direction at B which is a certain distance and direction from an initial point A. Thus, by randomly

choosing a point A, and selecting an angle, α and considering a point a distance k σ away (again $\sigma = 4$ is used as the resolution parameter of the first order derivative Gaussian operator), to find B, see Figure 7.12. Then define, the angle β to be the gradient direction at B, relative to the gradient direction at A.

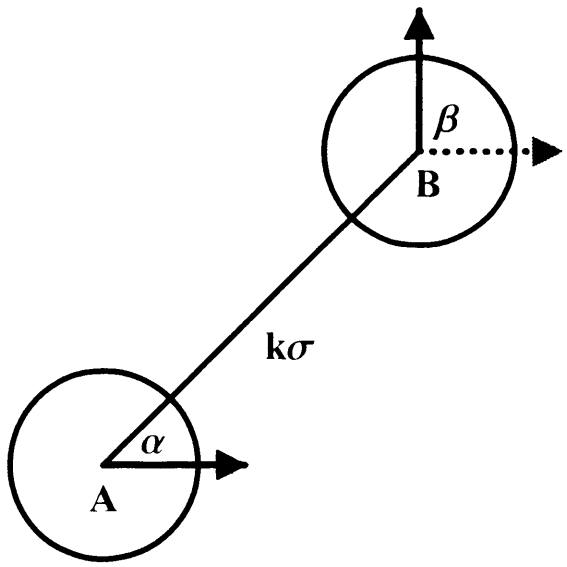


Figure 7.12: Illustrates an example of data collection to determine how the dependencies between two gradient directions changes depending on their relative orientation. In the diagram, a point A is chosen and its gradient direction (depicted by solid black arrow at A) is computed. Observing in a direction α and distance $k\sigma$ point B is located. Compute the the relative gradient direction β at B (the absolute gradient direction is shown as a dotted black arrow at B) as the difference between the two gradient directions at A and B.

7.2.1 Methodology: Information Contour Maps

In the data collection process, $10,000$ pairs of points are collected such as A and B (shown in Fig. 7.12) per image for 100 images whereby the initial location of the first point in each pair is chosen randomly. Again, the ensemble of natural images are taken from the van Hateren database [112] used earlier in the main results of the thesis; the composition of phase randomized natural images and Gaussian noise images is also as described in Section 5.1. For each image and set of $10,000$ points, a range of integer values of k are considered in steps of 1 in the range $[1, 16]$; and values of α in steps of $\pi/8$ in the range $(-\pi, \pi]$. The gradient directions measured at each location (at B it is the relative gradient direction compared to the gradient direction at A) and are recorded in bins of width $\pi/32$ over the range $(-\pi, \pi]$. Thus a 2-D probability distribution is obtained containing different separations between gradient direction measurements and different angles between the gradient directions: $p_k(\beta|\alpha)$.

Figure 7.13 illustrates the spatial extent of a first order derivative Gaussian operator of $\sigma = 4$ at A (yellow circle), and another operator, 10σ away at B (red circle), to give a sense of the separation of operators, relative to their size at these larger distances compared to the previous results which only considered separations up to $k = 5$.

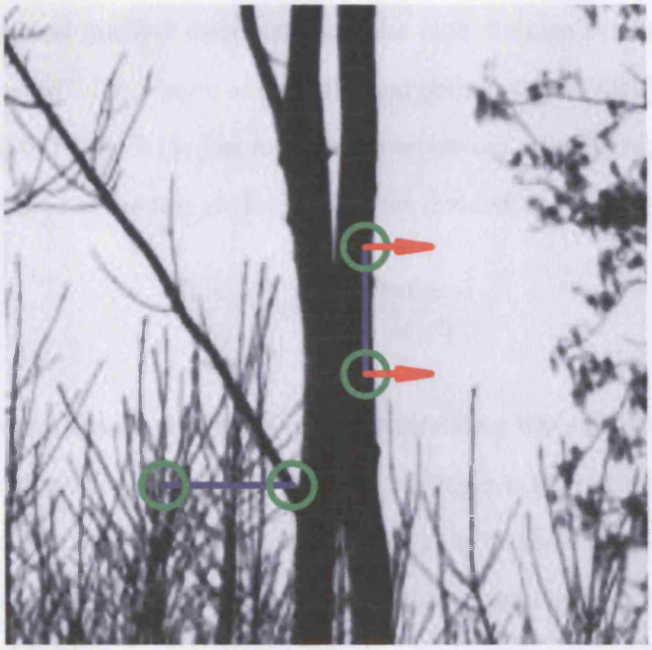


Figure 7.13: Illustrates the construction in Figure 7.12 (to find how 2-point gradient direction dependencies change with orientation) overlayed on a part of a natural image. Circles represent the width of two Gaussian operators 10σ apart, red circle and yellow circle. The circle radii are $\sqrt{3}\sigma$.

Information-Theoretic Calculations

In order to calculate the distribution of $p(\beta)$ where there is no knowledge of which direction α is chosen, and no knowledge of how far B is from A, the method described previously is repeated with randomly chosen B as well as A. Thus, using the histograms $p_k(\beta|\alpha)$ and $p(\beta)$, what information is gained from the gradient direction at B given the location of and gradient direction at A can be calculated. Again information theory is used to calculate the entropies of $p_k(\beta)$ and $p_k(\beta|\alpha)$ as shown in Equations 7.5 and 7.6.

$$H(\beta) = \sum_{\beta} -p(\beta).\ln[p(\beta)] \tag{7.5}$$

$$H_k(\beta|\alpha) = \sum_{\beta} -p_k(\beta|\alpha).\ln[p_k(\beta|\alpha)] \tag{7.6}$$

Thus, to compute the information obtained from knowledge of k and α, the mutual information is computed:

$$I_k(\beta;\alpha) = H(\beta) - H_k(\beta|\alpha). \tag{7.7}$$

The results are expressed using 2-D contour plots that depict how the amount of information between a second gradient direction decreases with distance and angle from an initial gradient direction located at the centre of each plot and pointing horizontally right, as shown in the bottom right plot of Figure 7.14. The information scores are actually fractions of the mutual information shared between the two gradient directions divided by $H(\beta)$:

$$\frac{I_k(\beta;\alpha)}{H(\beta)} = 1 - \frac{H_k(\beta|\alpha)}{H(\beta)} \tag{7.8}$$

Equation 7.8 represents the fraction of shared information which is obtained from knowledge of an initial gradient direction and the distance and angle to the second gradient direction. For instance, if no information is gained from knowledge of α, then Equation 7.8 would evaluate to zero. In essence both equations (Eqn. 7.7) describe the information gained about the gradient direction of B from A, given knowledge of α for specific separations k.

7.2.2 Results: Information Contour Maps

The results in Figure 7.14 show that for ensembles of natural images there is more dependency at locations perpendicular to the gradient direction at A than for parallel directions. The anisotropy can be quantified with the following rule: the same fractional amount of information is retained almost twice as far perpendicular than parallel. For $k < 5$ there is very little difference in the contours between natural random phases and natural images. Moreover, it should be noted that the contours are determined from a logarithmic scale thereby emphasising small information changes especially for $k > 5$ where differences emerge between natural images and phase randomized natural images.

Gaussian noise images, as expected, show marked differences from natural or phase randomized natural images; dependency drops to 1/100 in all directions by $k = 3$. There is also a slight bias toward information being retained parallel to the initial gradient direction rather than perpendicular, which is in contrast to natural images.

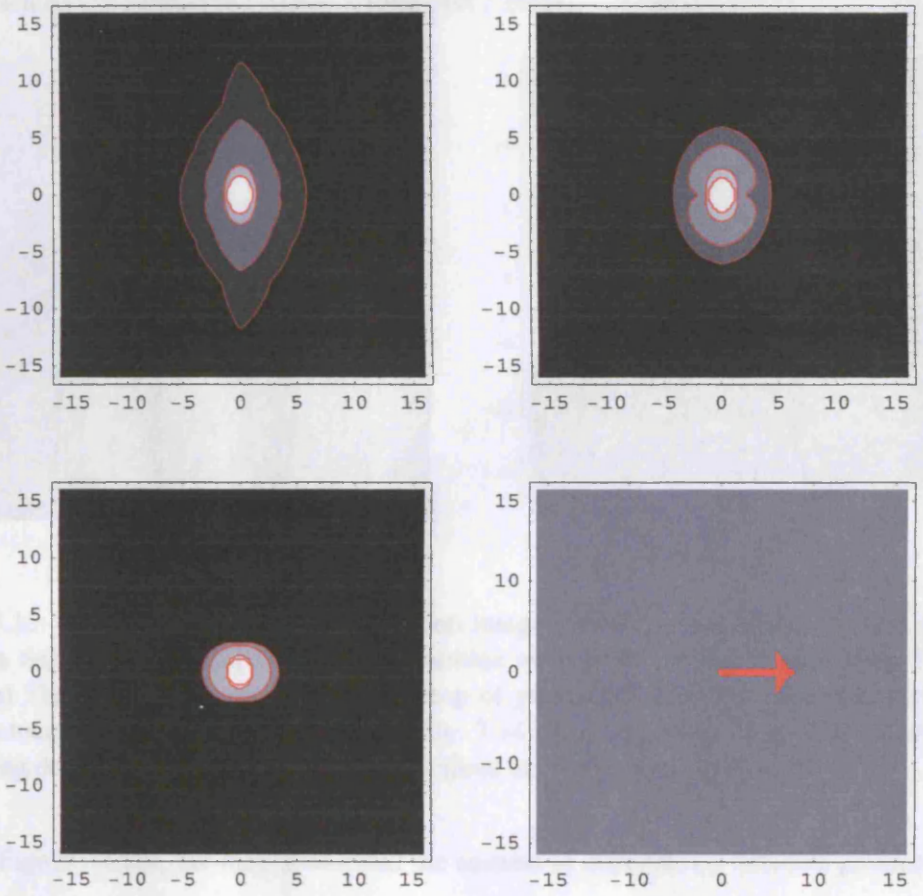
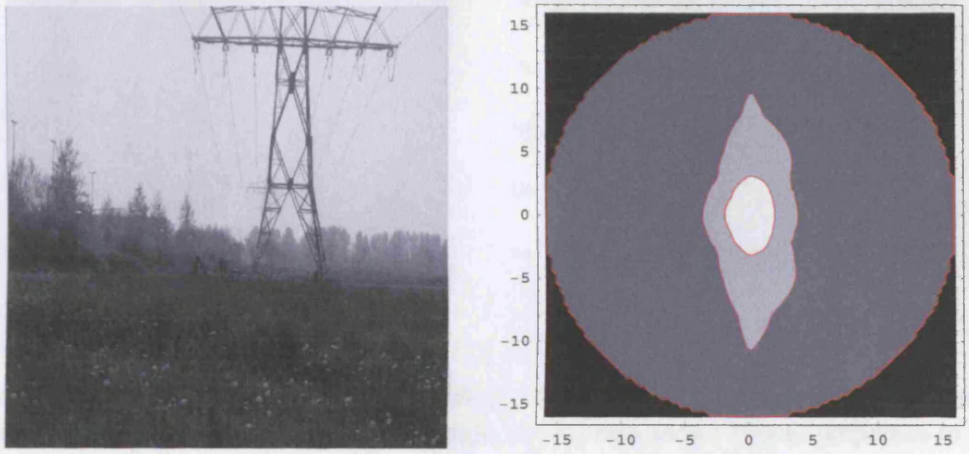


Figure 7.14: Information contour maps of gradient direction dependencies for three different image classes: natural (top left), phase randomized natural (top right), and Gaussian noise (bottom left). The centre of each map is the position of a gradient direction pointing horizontally to the right as depicted in bottom right map. Red contours are fractions of information at $1/10, 1/20, 1/100, 1/200$ as described by Eqn. 7.8 on a logarithmic scale. For $\frac{H_k(\beta|\alpha)}{H(\beta)} = 1$, $\frac{I_k(\beta;\alpha)}{H(\beta)}$ evaluates to zero and is plotted here as black. Lighter shading corresponds to higher fractional information. The horizontal and vertical axes represent the values of k (separation between the pairs of gradient directions). Therefore, moving away from the centre—the first gradient direction location—the gradient direction dependency between gradient directions is shown to decrease radially for all image classes.

Individual Natural Images

It is interesting to examine features in individual natural images that give rise to information contour maps which have either markedly increased dependency or decreased dependency compared to the ensemble average of natural images (Fig. 7.14 top left). Such examples are manually chosen and illustrated in Figures 7.15(a) and 7.16(a).

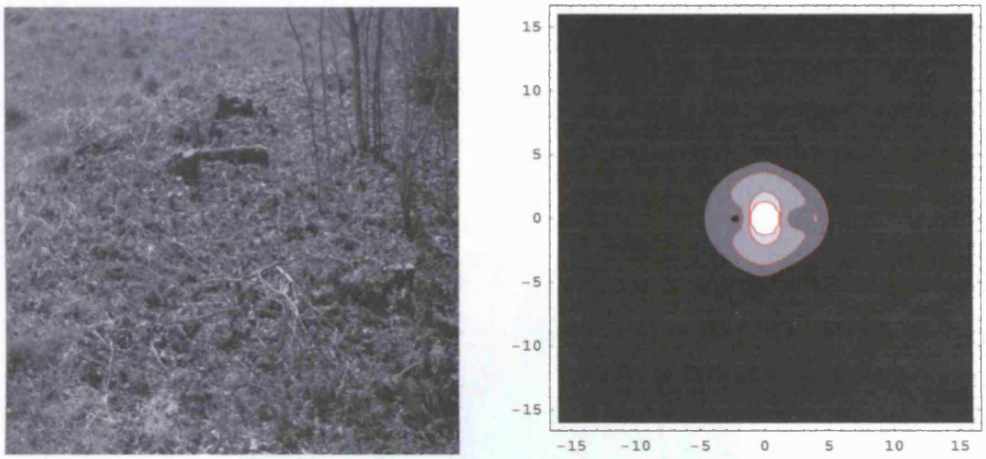


(a) (b)

Figure 7.15: (a) An example of a single natural image which displays longer 2-point gradient direction dependencies compared to the ensemble average of natural images (Fig. 7.14 top left). (b) The 2-point information contour map of gradient direction dependencies of image (a). Contours here are the same as used in Fig. 7.14. The outermost contour is an artifact of measuring dependency between gradient directions only to a radius of $k = 16$.

In Figure 7.15(a), the map shows that the amount of dependency between gradient directions is retained for greater distances compared to the mean results for an ensemble of natural images. The physical cause of the increased dependency between gradient directions is most likely a result of the 'smoothness' of the sky. In contrast, Figure 7.16(a)) illustrates a natural image whose information contour map (Fig. 7.16(b)) resembles that of the ensemble average for phase randomized images (Fig. 7.14 top right). This is not surprising as Figure 7.16(a)) is a scene of vegetation which appears more 'noise-like'; there is little sense of structure in the image.

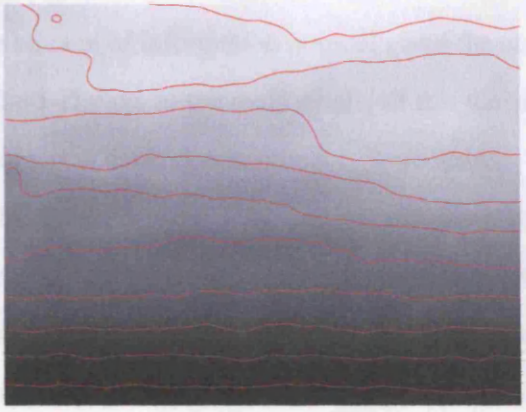
The 'smoothness' feature of sky is illustrated for the natural image in Figure 7.18(a). The contour lines in red show the homogeneity across the sky in terms of intensity. However, in the ground section (fig. 7.17), a different pattern emerges, which tends to point to the fact, sky is the more significant part of the image contributing to the extended gradient direction dependencies in Figure 7.15(a).



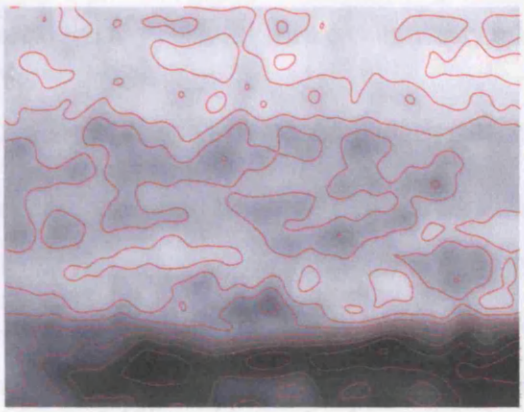
(a) (b)

Figure 7.16: In (a) is an example of a natural image. Illustrated in (b) is the information contour map of gradient direction dependencies of image (a); the map shows similar properties to that of an ensemble of phase randomized images (shown in Fig. 7.14 top right).

Figure 7.17: Image of vast clear sky and dark ground



(a)



(b)

Figure 7.18: Both images blurred to scale $\sigma = 8$. Contour lines in red, represent regions of equal intensity. Shown in (a) is the sky part of Fig.7.17 and in (b) the ground part.

7.2.3 Conclusions: Information Contour Maps

The 2-point dependency maps show that differences do emerge between natural images and phase randomized images. Therefore, it might be argued that the results for 2-point gradient direction dependency are inconsistent with the results in Section 6.2 which show that the dependency between gradient directions of an ensemble of natural images is determined by its mean power spectrum. However, the dependency shown in the contour maps are a function of distance and angle, not just distance. Further, the difference which emerge here are only for distances of $k > 5$ where the amount of information is small given the contours in Figure 7.14 are on a logarithmic scale. The 1-D plots in the main results of this thesis (Fig. 6.6), also indicate that the 2-point gradient direction dependencies for $k > 5$ are weak.

Interestingly, the information contour maps of gradient direction dependency in natural images are consistent with other studies by Sigman et al. [9] and Geisler et al. [180], which indicate that an edge element is most likely to be co-linear to a reference edge element. This is supported here by the fact that given an initial gradient direction, dependency with another gradient direction extends furthest in a perpendicular direction to the initial gradient direction. However, the results do not exhibit any evidence for a co-circularity rule for gradient directions.

7.3 Experiments to find the Source of Synergetic Dependencies

In this section, experiments are conducted to provide insight into the possible sources of the synergetic dependencies found between gradient directions in the main results of this thesis as described in Section 6.4. The first experiment is to recompute the 3-point gradient direction dependencies for a line, rather than a triangular, configuration of measurements. The second experiment involves computing dependencies of gradient vector fields with non-zero curl. This is because of a mathematical property that the gradient of a scalar field (e.g. images) has zero curl, and it is hypothesised that this may act as an additional constraint on the dependencies of gradient directions in images. Details are also provided on additional experiments looking at other random distributions of noise images and the search for significant configurations of gradient directions.

7.3.1 Line Configuration

An inherent problem in computing the 3-point gradient direction dependencies for a line spatial configuration is that two different distances exist between the three points. Given three points

A,B,C in a line, there will be two pairs of points equidistance from one another, say A—B and B—C, but also one other pair A—C twice the separation of the other pairs. The horizontal axis in Figure 7.19 shows the separation between measurements k, as the shorter distance. This makes the comparison of 3-point gradient direction dependencies for a line configuration with the equilateral triangular configuration not particularly straightforward; although, comparison is somewhat aided by the fact the 3-point interaction information, given by $I(A;B;C)$, is symmetric with respect to A, B, and C unlike, for example, $I(A;B|C)$.

7.3.2 Results and Discussion: Line Configuration

Figure 7.19 shows that the 3-point gradient direction dependencies for image classes with the same power spectrum, i.e. natural images and phase randomized natural images, and whitened natural images and Gaussian noise images, the gradient direction dependencies are still indistinguishable even for a line spatial configuration, and in this respect the results herein support the 3-point gradient direction dependencies found for an equilateral triangular spatial configuration of measurements.

However, the 3-point dependencies between gradient directions in a line spatial configuration show no synergy (synergy would be indicated by negative values in Fig. 7.19) for natural images and phase randomized images, unlike the triangular configuration results shown in Figure 6.10. But, synergy is found in whitened natural images and Gaussian noise images, as shown in Figure 7.19. It is noticeable that, for Gaussian noise images, the maximum amount of synergy occurs at half the k value compared with the results from the 3-point gradient direction dependencies in Figure 6.10 (top) although this might in fact be due to the existence of two unique distances—one half the distance of the other—between the three gradient directions in a line spatial configuration.

7.3.3 Curl of a Gradient Field

In this section, it is investigated whether the mathematical property that the gradient of a scalar field has zero curl constrains the dependency between gradient directions.

Helmholtz's theorem states that a vector field can be fully determined by its curl and divergence. The curl describes the rotation in a small region of the field, and div describes the flow

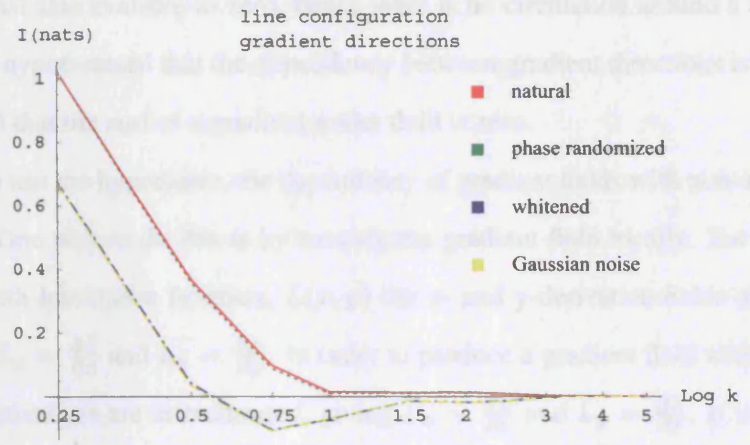


Figure 7.19: Illustrates results of the interaction information for 3-point line configuration of gradient directions. Compare this figure with Figure 6.10 which is for a triangular configuration of gradient directions. Vertical axis: $I(A;B;C)$ interaction information (nats); horizontal axis: distance (logarithmic scale) between measurements in multiples k of σ. Image key: natural (red solid), phase randomized natural (green dotted), whitened natural (blue dash-dot), Gaussian noise (yellow dashes). Recall, positive (negative) values of information indicate dependencies between the triple of gradient directions which are redundant (synergetic).

into or out from a surface. For the gradient of a scalar field it can be shown to have zero curl:

$$\mathbf{F}(x,y) = \nabla\phi \tag{7.9}$$

where, ∇ is the vector differential operator $\{\mathbf{i}\,\partial_x, \mathbf{j}\,\partial_y\}$, where $\partial_\psi = \frac{\partial}{\partial\psi}$, and if ϕ is a 2-D scalar field then curl($\mathbf{F}$) is given by:

$$\nabla \times \mathbf{F} = (\partial_y\partial_x\phi - \partial_x\partial_y\phi)\,\mathbf{k} = 0 \tag{7.10}$$

where $\mathbf{k} = \mathbf{i} \times \mathbf{j}$. Thus, gradient scalar fields are not unrestricted vector fields.

The restriction on rotation in a small region (zero curl) also restricts the rotation in larger regions (zero circulation) due to the vector identity, known as Green's theorem, which states that the line integral of a vector field $\mathbf{F}$ around a closed curve C is equal to the surface integral over the region bounded by C (denoted D) of the curl of the vector field:

$$\int_C \mathbf{F}.d\mathbf{s} = \int\int_D \nabla \times \mathbf{F}\, dA. \tag{7.11}$$

Physically, Green's theorem amounts to: the circulation around curve C is equal to the sum of all circulations in a small region inside C. Therefore, for gradient scalar field ($\mathbf{F} = \nabla\phi$) the

line integral must also evaluate to zero, hence there is no circulation around a closed curve in $\nabla\phi$. Thus, it is hypothesised that the dependency between gradient directions is constrained by the requirement that the curl of a gradient scalar field is zero.

In order to test the hypothesis, the dependency of gradient fields with non-zero curl should be computed. One way to do this is by rotating the gradient field locally. For example, for a given image with luminance function, $L(x, y)$ the x- and y-derivative fields are given by L_x and L_y where $L_x = \frac{\partial L}{\partial x}$ and $L_y = \frac{\partial L}{\partial y}$. In order to produce a gradient field with non-zero curl, the x- and y-derivatives are interchanged, giving $L_x = \frac{\partial L}{\partial y}$ and $L_y = \frac{\partial L}{\partial x}$. It is important that when the curl of the vector field is made non-zero, the div is not then made to be zero because this also restricts the vector field, and therefore the div must also be non-zero. Therefore, in this experiment, the 3-point gradient direction dependencies in a triangular and line spatial configuration are computed from gradient vector fields manipulated to have non-zero curl and non-zero divergence ($\nabla \times \nabla\phi \neq 0$, $\nabla.\nabla\phi \neq 0$). These computations are to be compared with results from the unmanipulated gradient vector fields which contain zero curl and non-zero divergence ($\nabla \times \nabla\phi = 0$, $\nabla.\nabla\phi \neq 0$), the results of which have been described in Sections 6.4 and 7.3.2.

7.3.4 Results and Discussion: Curl of Gradient Field

Figure 7.20 top shows 3-point gradient direction dependencies $I(A; B; C)$ for a triangular configuration points, after the gradient vector field has been altered to have both non-zero curl and non-zero divergence ($\nabla \times \nabla\phi \neq 0$, $\nabla.\nabla\phi \neq 0$); this chart should be compared to the chart appearing in Section 6.2. The most obvious difference between the charts is that the dependency for a non-zero curl gradient field (Fig. 7.20 top) has no synergetic dependencies. In fact, overall, the redundancy between triples of gradient directions has increased, and the point at which synergy was a maximum for natural images is now the point at which the dependency becomes zero at $k = 2.5$, and similarly for Gaussian noise images at $k = 1.5$. The increase in redundancy implies that, in comparison to the curl-free vector field, the mutual information between the gradient directions A and B when C is known is less. This is to be expected because by making the vector field have non-zero curl, the gradient directions are less constrained, and thereby there is less dependency between gradient directions.

Figure 7.20 bottom shows the 3-point gradient direction dependencies for a line configuration computed from a gradient vector field with non-zero curl and non-zero div. The de-

pendencies for a given k are similar to that found in Figure 7.19, which were obtained from an unmanipulated gradient vector field. Therefore, the curl-free property is not a constraint between gradient directions in a line configuration. This is not surprising, notwithstanding that the circulation of such a field is zero, because the line integral is over a closed curve. When computing the dependencies for a line configuration we have effectively an open curve. Therefore we might not expect the curl-free property of the gradient vector field to be a constraint in a line configuration.

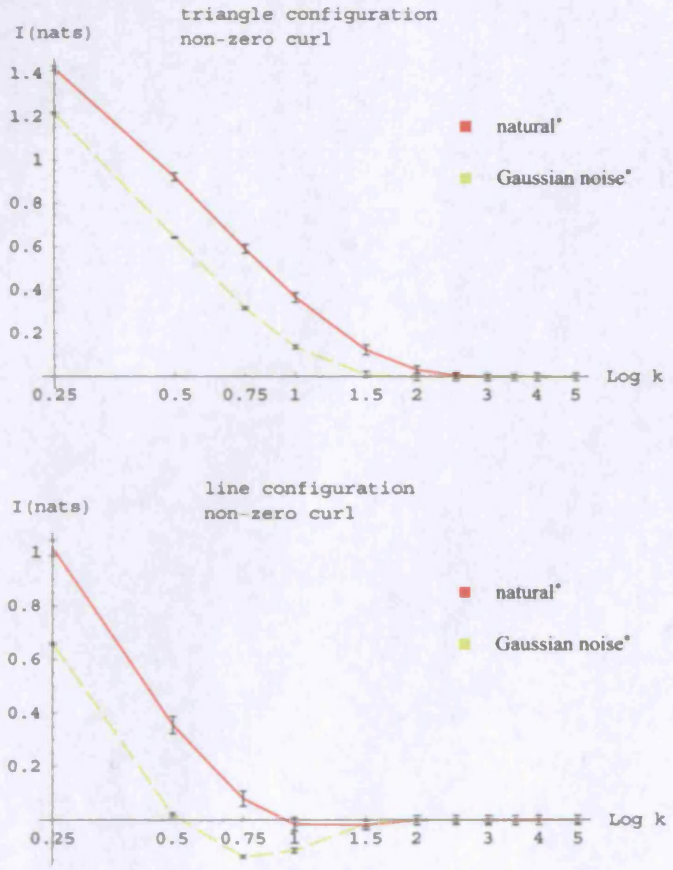


Figure 7.20: Shows 3-point gradient direction dependencies $I(A;B;C)$ of gradient vector fields manipulated from natural and Gaussian noise images to have non-zero curl and non-zero divergence, for triangular configuration (top) and line configuration (bottom). Top chart should be compared to Fig. 6.10 top, and bottom chart to Fig. 7.19. Vertical axis: $I(A;B;C)$ interaction information (nats); horizontal axis: distance (logarithmic scale) between measurements in multiples k of σ. *Red solid curves are results from the manipulated gradient vector field of natural images, and yellow dashes from Gaussian noise images.

7.3.5 Other Random Distributions

The observation that Gaussian noise images have synergetic dependencies between gradient directions in both triangular (Fig. 6.10) and line (Fig. 7.19) configurations, led to the hypothesis that the mean-reverting property of drawing pixel values from a Normal distribution contributes to the dependency between gradient directions. Mean reversion here is the tendency of luminance values to revert to a mean value, which is the mean of the Normal probability distribution (from which the pixel values are drawn).

To examine this hypothesis, the luminance functions of images generated from two other random distributions which have higher and lower kurtosis than the Normal distribution were also examined. The distributions studied included the Laplace distribution: fatter tails than a Normal distribution, therefore higher kurtosis and a greater chance of high and low pixel values; and the Uniform distribution: less kurtosis than a Normal distribution with no tendency for any mean value.

Examination of the results showed no significant differences in gradient direction dependencies between images generated from these distributions and Gaussian noise images.

7.4 Gradient Direction Configurations

Attempts were made to find if certain configurations of gradient directions contributed substantially more information to the overall dependency between 3-point measurements. Such an approach would provide qualitative differences between the image classes by comparing which gradient direction configurations were most dependent in each image class and which configurations provided synergetic dependencies. For example, do the more dependent gradient direction configurations happen to be symmetrical in ensembles of natural images?

Investigations of all the image classes suggest that, although certain configurations dominate, no obvious pattern (such as symmetry or configurations corresponding to obvious image structures) in the top ten most dependent configurations were found. This is also the case even after pooling over configurations that are transformations in terms of rotation and reflection, an example of such a configuration group is shown in Figure 7.21.

However, if we simply search for the most frequently occurring configurations in each image class, more obvious patterns emerge. Figure 7.22 shows an example taken from configurations where $k = 2$ (see Fig. 5.10). The results at $k = 2$ are not necessarily followed at

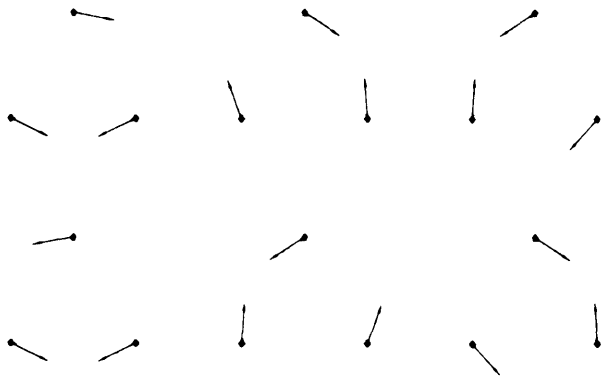


Figure 7.21: Top row shows 3-point gradient direction configurations which are all equivalent under a $\frac{2\pi}{3}$ rotation of each triangle. Bottom row shows the reflections of the top left triangle in vertical plane, and its rotations.

other values of k, but interestingly, for $k = 2$ (see Fig. 5.10), the synergetic dependencies are equivalent for all four image classes, as shown in Figure 6.10.

For both natural images and phase randomized natural images, the most frequently occurring configuration at $k = 2$ is where the gradient directions are all, approximately, in the same direction. For whitened natural images and Gaussian images this is not the case, the most frequently occurring configuration is that of light and dark blobs, where the gradient flows into or out from the centre of the configuration.

Despite all four image classes having the same synergetic dependencies, there are qualitative differences between image classes in terms of the most frequent gradient direction configurations found, although we find that natural and phase randomized natural images are similar, as are whitened and Gaussian noise images at $k = 2$. This similarity between the pairs of image classes holds for $k \leq 2$. For $k \leq 1$, all four image classes have gradient direction configurations which tend to be in the same direction, as illustrated in the top two rows of Figure 7.22.

Interestingly, for all four image classes at $k = 2$, where synergetic dependencies exist, the least frequently occurring configurations are when the gradient directions indicate circulation in the gradient field, as shown in Figure 7.23 (example is for natural images). This observation bears some consistency with the notion that the zero-curl property of the gradient of a scalar field contributes to synergetic dependencies, which was discussed in Section 7.3.3. Although, similar configuration patterns are also found for $1 < k \leq 5$ for all image classes. Recall that no synergy is found for $k \leq 1$.

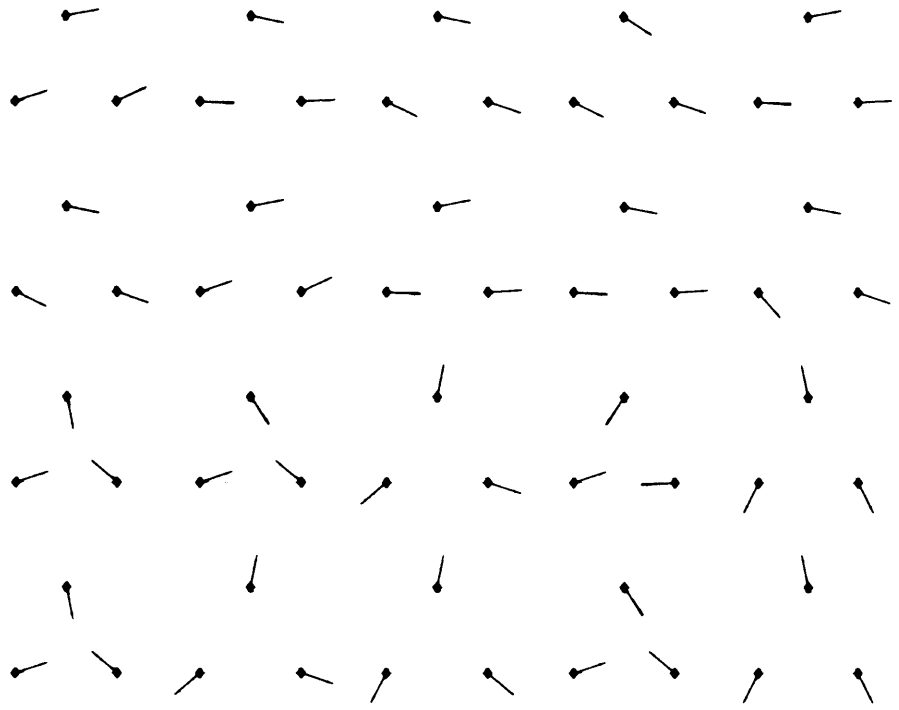


Figure 7.22: Most frequently occurring 3-point gradient direction configurations for separation $k = 2$, (see Fig. 5.10 for details of k) in the following ensembles of images, from top row to bottom row: natural, phase randomized natural, whitened natural, and Gaussian noise.

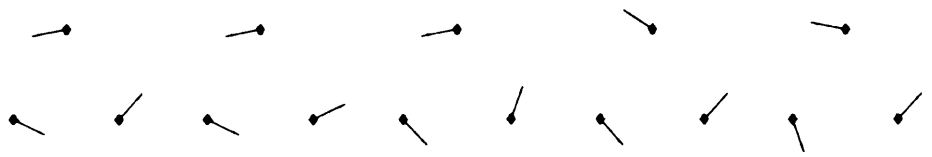


Figure 7.23: Least frequently occurring 3-point gradient direction configurations for separation $k = 2$, (see Fig. 5.10 for details of k) for an ensemble of natural images, and this is also the case for phase randomized natural images, whitened natural images, and Gaussian noise images.

7.5 Conclusions of Further Experiments

The main results in the previous chapter, Section 6.5, show that average dependencies between 2-point and 3-point gradient directions of an ensemble of natural images are strongly dependent upon the ensemble's mean power spectrum. This conclusion was primarily based on the results which show the amount of 2-point and 3-point gradient direction dependency is the same in natural images and phase randomized natural images; and also the same for whitened natural images and Gaussian noise images.

In this chapter, it has been shown that ensembles of images with varying power law power spectra and random phases, show an approximate lawful relationship between the power law exponent of the ensemble's mean power spectrum and the dependency between nearby gradient directions. But this lawful relationship breaks down for ensembles of scale-variant ensembles such as images generated from constant-sized disks. If, the dependency between gradient directions is computed for an ensemble of dead leaves model images, which are a model of images that reproduce the scale-invariant properties of natural images such as the $1/f^2$ ensemble power spectrum, then the dependencies are similar to an ensemble of natural images.

In this chapter it has also been demonstrated that the 3-point dependencies of gradient directions at multiple locations does, to some extent, depend on the spatial configuration of those locations. This is based on comparing the 3-point statistics of gradient directions in an equilateral triangular spatial configuration and a line spatial configuration. The triangular configuration, shows synergetic dependencies for all four image classes (natural, phase randomized natural, whitened, Gaussian noise), however, for the line configuration, only whitened natural images and Gaussian noise images contain synergetic dependencies, previously in a triangular configuration all four image ensembles had synergetic dependencies (Sec. 6.4). Recall synergetic dependencies indicate that the mutual information between two gradient directions, A and B is greater when a third gradient direction is known than when it is not known.

Investigations of the possible causes of synergy observed in the triangular configuration of triples of gradient directions show that it is likely to be a result of the mathematical property that the gradient of a scalar field has zero curl. However, this restriction on the curl of the field is not relevant for a line configuration of gradient directions because it is not around a closed curve. To summarise these results, Table 7.1 shows the configuration of measurements, whether the gradient vector field is curl-free or not, and which image classes contain synergetic gradient

direction dependencies.

	Triangular	Line
Curl= 0	N, P, W, G	W, G
Curl≠ 0	None	W, G

Table 7.1: Summarises under what circumstances the four image classes: natural (N), phase randomized (P), whitened natural (W), and Gaussian noise (G) have synergetic gradient direction dependencies. Top row indicates the configuration of measurements, and first column indicates whether the gradient vector field has been manipulated to have non-zero curl.

Chapter 8

Preliminary Future Work and Conclusions

In Section 8.1 of this chapter, the methodology and results of preliminary future work are presented and the significance of the findings are discussed in relation to the main results of this thesis. In Section 8.2, a summary of the thesis is given together with the main contributions of this thesis.

8.1 Preliminary Future Work

The main work reported in this thesis could be extended in two ways: i) computing n-point dependencies for $n > 3$ (i.e. for more than triples of measurements), and for ii) computing dependencies between quantities depending on second order derivative measurements, for example the shape index. It turns out that it is most convenient, from a computational viewpoint, to do i) and ii) simultaneously. This is because of constraints on performing calculations on the probability distributions. Previously, the statistics on gradient directions, which ranged over $[-\pi, \pi)$, were collected into 16 bins of width $\pi/8$. This meant when forming the histograms for 3-point dependency, it was necessary to populate $16^3 = 4096$ bins. Extending this to the dependencies between five measurements would make it necessary to populate over one million bins ($16^5 = 1,048,576$), which can lead to greater errors in calculating the entropy estimates from these histograms if the bins are not sufficiently populated. To compute 9-point dependencies would require almost seventy billion bins meaning a considerable amount of data would need to be collected which would not be manageable. The 9-point dependency is interesting because

in a 3×3 pixels configuration the centre pixel is completely surrounded as illustrated by:

$$\begin{bmatrix} X_5 & X_1 & X_6 \\ X_2 & X_C & X_3 \\ X_7 & X_4 & X_8 \end{bmatrix}. \tag{8.1}$$

A solution to the problem of increased dimensionality is to reduce the number of categories of gradient directions to just five. With five categories the number of bins which must be populated for 5-point dependencies is just over three thousand $5^5 = 3,125$; for 9-point dependencies it is close to two million, which is still manageable. However, there is no intuitive or reasonable way to split gradient directions into five categories. This is a reason for moving to second order derivative measurements and the shape index; it is possible to split the shape index into five geometrically-meaningful categories.

The shape index describes the curvature of a point in the image, as explained in Section 2.4.5, but this will be briefly summarised again here. Second order derivative measurements are given by the Hessian matrix, which calculates the gradient of the gradient, i.e. the curvature of a point not the slope:

$$\text{Hessian } L(x,y) = \begin{pmatrix} \frac{\partial^2 L}{\partial x^2} & \frac{\partial^2 L}{\partial x \partial y} \\ \frac{\partial^2 L}{\partial y \partial x} & \frac{\partial^2 L}{\partial y^2} \end{pmatrix}. \tag{8.2}$$

It is possible to define curvature at a point in any direction but there are just two principle curvatures κ_1, κ_2 (maximal and minimal curvature) found from the eigenvectors of the Hessian. This leads to the shape index, which is computed from an arbitrary constant multiplied by the arctangent of a specific ratio of κ_1, κ_2:

$$\text{Shape Index} = \frac{2}{\pi} arctan\left(\frac{\kappa_1 + \kappa_2}{\kappa_1 - \kappa_2}\right), \kappa_1 \leq \kappa_2 \tag{8.3}$$

and the magnitude of curvature is given by the curvedness:

$$\mathrm{c} = \sqrt{\frac{\kappa_1^2 + \kappa_2^2}{2}} \tag{8.4}$$

The shape index is often split into nine categories or labels: cup, trough, rut, saddle rut, saddle, saddle ridge, ridge, dome, and cap. However, another choice is to break the shape index into

just five categories with intervals across the range $[-1, 1)$ of:

- cup:$[-1, -13/16)$,
- rut: $[-13/16, -6/16)$,
- saddle: $[-6/16, 6/16)$,
- ridge: $[6/16, 13/16)$,
- cap: $[13/16, 1)$.

The corresponding shapes to these five categories are illustrated in Figure 8.1.

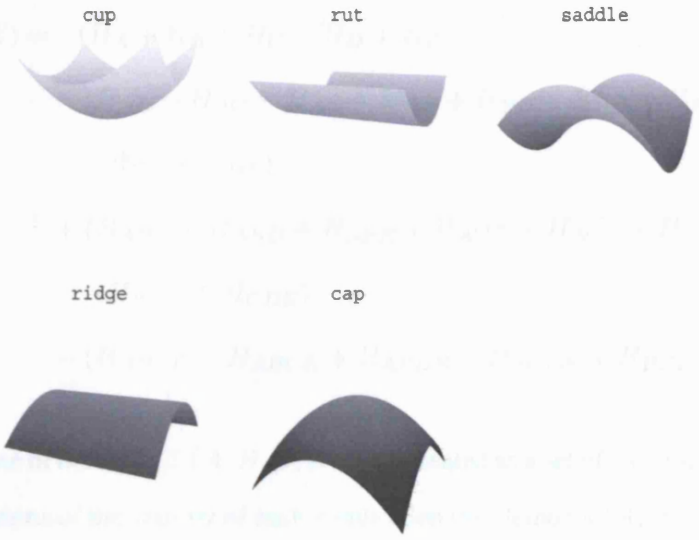


Figure 8.1: A convenient progression from 2-point and 3-point gradient direction dependencies is to study a greater number of points and 2nd-order derivative measurements (e.g. shape index). The shape index (derived from the curvature of image points) can be divided into the five categories illustrated above. Having just five categories makes it computationally feasible to compute the dependencies in 9-point configurations.

8.1.1 Extending Information Theory to 9-point Dependencies

Previously, in Section 4.1.3 the 3-point interaction information $I(A;B;C)$, and total 3-point mutual information $I(A,B,C)$ were expressed in terms of joint and marginal entropies as:

$$I(A;B;C) = H(A)+H(B)+H(C)-(H(A,B)+H(A,C)+H(B,C))+H(A,B,C) \quad (8.5)$$

and,

$$I(A,B,C) = H(A) + H(B) + H(C) - H(A,B,C). \quad (8.6)$$

Recall that interaction information of an n-point configuration excludes any $(n-1,2,\ldots)$-point configuration dependencies. Thus, extending this to 4-point interactions where for clarity $H(X,Y)$ is denoted by H_{XY}:

$$\begin{aligned} I(A;B;C;D) = \; & (H_A + H_B + H_C + H_D) \\ & - (H_{AB} + H_{AC} + H_{AD} + H_{BC} + H_{BD} + H_{CD}) \\ & + (H_{ABC} + H_{ABD} + H_{ACD} + H_{BCD}) - H_{ABCD}. \end{aligned} \tag{8.7}$$

and to 5-point interactions:

$$\begin{aligned} I(A;B;C;D;E) = \; & (H_A + H_B + H_C + H_D + H_E) \\ & - (H_{AB} + H_{AC} + H_{AD} + H_{AE} + H_{BC} + H_{BD} + H_{BE} + H_{CD} + \\ & \quad H_{CE} + H_{DE}) \\ & + (H_{ABC} + H_{ABD} + H_{ABE} + H_{ACD} + H_{ACE} + H_{BCD} + H_{BCE} + \\ & \quad H_{BDE} + H_{CDE}) \\ & - (H_{ABCD} + H_{ABCE} + H_{ABDE} + H_{ACDE} + H_{BCDE}) + H_{ABCDE}. \end{aligned} \tag{8.8}$$

Therefore, for ease of notation, if $\{A, B, C, D, E\}$ is treated as a set of 5 elements, Equation 8.8 can be recast in terms of the entropy of each k-subset on the elements $\{A, B, C, D, E\}$, to give:

$$\begin{aligned} I(A;B;C;D;E) = \; & H\{\text{1-Subset}_{ABCDE}\} - H\{\text{2-Subset}_{ABCDE}\} \\ & + H\{\text{3-Subset}_{ABCDE}\} - H\{\text{4-Subset}_{ABCDE}\} \\ & + H\{\text{5-Subset}_{ABCDE}\} \end{aligned} \tag{8.9}$$

where, for example,

$$(H_{ABCD} + H_{ABCE} + H_{ABDE} + H_{ACDE} + H_{BCDE}) \equiv H\{\text{4-Subset}_{ABCDE}\}.$$

Note that the number of k-subsets from a set of r elements is given by the binomial coefficient $\binom{r}{k} \equiv \frac{r!}{k!(r-k)!}$. Extending Equation 8.9 to 9 elements $\{A, B, C, D, E, F, G, H, I\}$, gives:

$$I(A;B;C;D;E;F;G;H;I) = \sum_{r=1}^{9}(-1)^{r-1}H\{\text{r-Subset}_{ABCDEFGHI}\} \tag{8.10}$$

and the total information is:

$$I(A,B,C,D,E,F,G,H,I) = H\{\text{1-Subset}_{ABCDE}\} - H(ABCDEFGHI). \tag{8.11}$$

8.1.2 Methodology for 9-point Analysis

The four image classes: natural, phase randomised natural, whitened natural and Gaussian noise (described in Section 5.1) are analysed for 9-point dependencies of shape index. Each image $L(x, y)$ is of dimensions 1024 by 1536, and each ensemble contains 100 images from each class. The distance between points is computed by altering the values of two parameters: σ and k. First, σ is the resolution of the Gaussian derivative operator $G(x, y; \sigma)$ used to extract the shape index $SI(x, y; \sigma)$ from the image $L(x, y)$. The shape index is given by:

$$SI(x, y; \sigma) = 2\arctan\left(\frac{L \otimes \partial_{xx}G + L \otimes \partial_{yy}G}{\sqrt{4(L \otimes \partial_{xy}G)^2 + (L \otimes \partial_{xx}G - L \otimes \partial_{yy}G)^2}}\right) \tag{8.12}$$

where for brevity $\partial_{xx}G \equiv \frac{\partial^2}{\partial x^2}G(x, y; \sigma)$ and $\otimes$ denotes convolution. Second, $SI(x, y; \sigma)$ is divided into partitions of size k and σ. The values of k observed are $\{0.25, 0.5, 0.75, 1, 1.5, 2\}$, and $\sigma = \{4, 2, \frac{4}{3}, 1, \frac{4}{3}, 1\}$, to give the following convenient pairs of value for $k\sigma$:

$$k\sigma = \begin{cases} 1 & \text{if } k = \{0.25, 0.5, 0.75, 1\} \\ 2 & \text{if } k = \{1.5, 2\}. \end{cases} \tag{8.13}$$

For example, to compute the shape index of a 6×6 image for $k\sigma = 2$ ($\sigma = 1$), the shape index of the image is computed and then partitioned into 2×2 subsets to form a 3×3 block such as,

$$\begin{bmatrix} X_{11} & X_{13} & X_{15} \\ X_{31} & X_{33} & X_{35} \\ X_{51} & X_{53} & X_{55} \end{bmatrix}$$

One complication which arises in this methodology is that the distance between each element X_{ij} is not fixed for any n-subset, i.e. there are several 5-point configurations in a given 3×3 pixel block. Thus, for any set of n-point configurations there are different distances between the elements, which is illustrated in Figure 8.2. Therefore, for each n-point configuration the average, minimum, and maximum dependency is computed.

$$\begin{bmatrix} X_{11} & & X_{15} \\ & X_{33} & \\ X_{51} & & X_{55} \end{bmatrix} \quad \begin{bmatrix} & X_{13} & \\ X_{31} & X_{33} & X_{35} \\ & X_{53} & \end{bmatrix} \quad \begin{bmatrix} & & X_{15} \\ X_{31} & X_{33} & X_{35} \\ & & X_{55} \end{bmatrix}$$

$$\begin{bmatrix} X_{11} & X_{13} & \\ & X_{33} & \\ & X_{53} & X_{55} \end{bmatrix} \quad \begin{bmatrix} X_{11} & & \\ & X_{33} & X_{35} \\ & X_{53} & X_{55} \end{bmatrix} \quad \begin{bmatrix} X_{11} & X_{13} & \\ X_{31} & & \\ X_{51} & X_{53} & \end{bmatrix}$$

Figure 8.2: Illustrates four (out of a possible 126) 5-point configurations in a given 3x3 pixel block. Different distances exist between elements X_{ij} in each configuration.

8.1.3 Preliminary Results

Preliminary results for the dependency of 9-point configurations of shape index are presented here. The results can be viewed from two important perspectives: i) how dependency varies with n for a given distance between points, as shown in Figure 8.3, and ii) how dependency falls with distance between points as shown in Figure 8.4; the second perspective is similar to the analysis of gradient direction dependencies computed in the main results of this thesis. Figure 8.3 shows that for all image classes the total dependency between n-point configurations increases with greater n. Observing the relationship in more detail, it is noticeable that for small k, the total dependency increases linearly with n. However, for larger separations between points, the total dependency grows supralinearly with n. Most significantly, the shape index dependency for natural images and phase randomized natural images are indistinguishable, as they are for whitened natural images and Gaussian noise images albeit lower. This is the same pattern which was observed for gradient direction dependencies in the main results of this thesis.

However, for larger values of k (i.e. $k \geq 1$), the shape index dependency becomes different between the image classes. The difference is most pronounced for $k = 2$, but surprisingly phase randomized natural images increase over natural images, as do Gaussian noise images over whitened natural images. Although, in all cases, the differences between the image classes is small such that if the results in Figure 8.3 are re-plotted in terms of how the dependency for each n-point configuration changes with different separations k, Figure 8.4 is obtained (this is

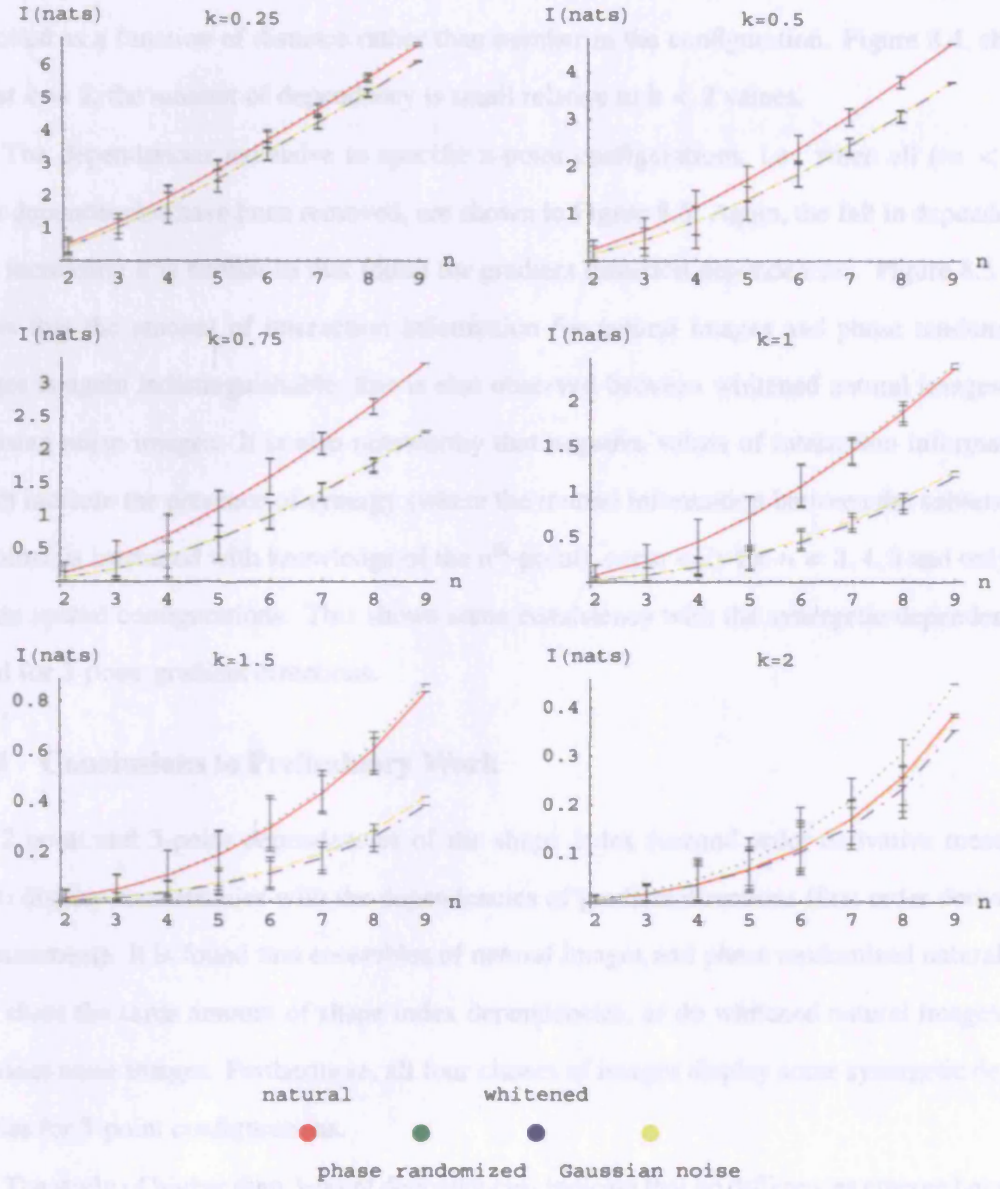


Figure 8.3: Illustrates how the shape index dependency between points varies with the number of points n in the configuration, at different distances between points k (see main text Sec. 8.1.2). Each curve in each chart represents different image classes. Vertical axis: the dependency between points in terms of total mutual information I(nats); horizontal axis: indicates the number of points in the configuration n. The error bars denote the maximum and minimum information scores for that n-point configuration.

the more familiar presentation of the results adopted for the 2-point and 3-point dependency of gradient directions). Figure 8.4, shows that natural and phase randomized image classes are indistinguishable, as are whitened natural and Gaussian noise image classes, where dependency is plotted as a function of distance rather than number in the configuration. Figure 8.4, shows that at $k = 2$, the amount of dependency is small relative to $k < 2$ values.

The dependencies exclusive to specific n-point configurations, i.e. when all $(m < n)$-point dependencies have been removed, are shown in Figure 8.5. Again, the fall in dependency with increasing k is similar to that found for gradient direction dependencies. Figure 8.5 also shows that the amount of interaction information for natural images and phase randomized images is again indistinguishable; this is also observed between whitened natural images and Gaussian noise images. It is also noteworthy that negative values of interaction information, which indicate the presence of synergy (where the mutual information between the subsets, (n-1)-points, is increased with knowledge of the n^{th}-point), occur only for $n = 3, 4, 5$ and only for certain spatial configurations. This shows some consistency with the synergetic dependencies found for 3-point gradient directions.

8.1.4 Conclusions to Preliminary Work

The 2-point and 3-point dependencies of the shape index (second order derivative measurement) display consistencies with the dependencies of gradient directions (first order derivative measurement). It is found that ensembles of natural images and phase randomized natural images share the same amount of shape index dependencies, as do whitened natural images and Gaussian noise images. Furthermore, all four classes of images display some synergetic dependencies for 3-point configurations.

The study of higher than 3-point dependencies indicate that no differences emerge between natural images and phase randomized images; for $k \leq 1.5$ this is also true between whitened natural images and Gaussian noise images, which both have less dependencies than natural images. For $k = 2$, small differences do emerge for shape index dependencies for all four image classes. Whether or not these results are significant will require further investigation because without an explanation for the difference, we cannot be satisfied that the effect is solely due to the properties of the image class. Although it is noteworthy that at $k = 2$ the amount of information is significantly smaller than for $k \leq 1$, and that for higher values of k there is less image data to populate the histogram bins on which the information scores are based.

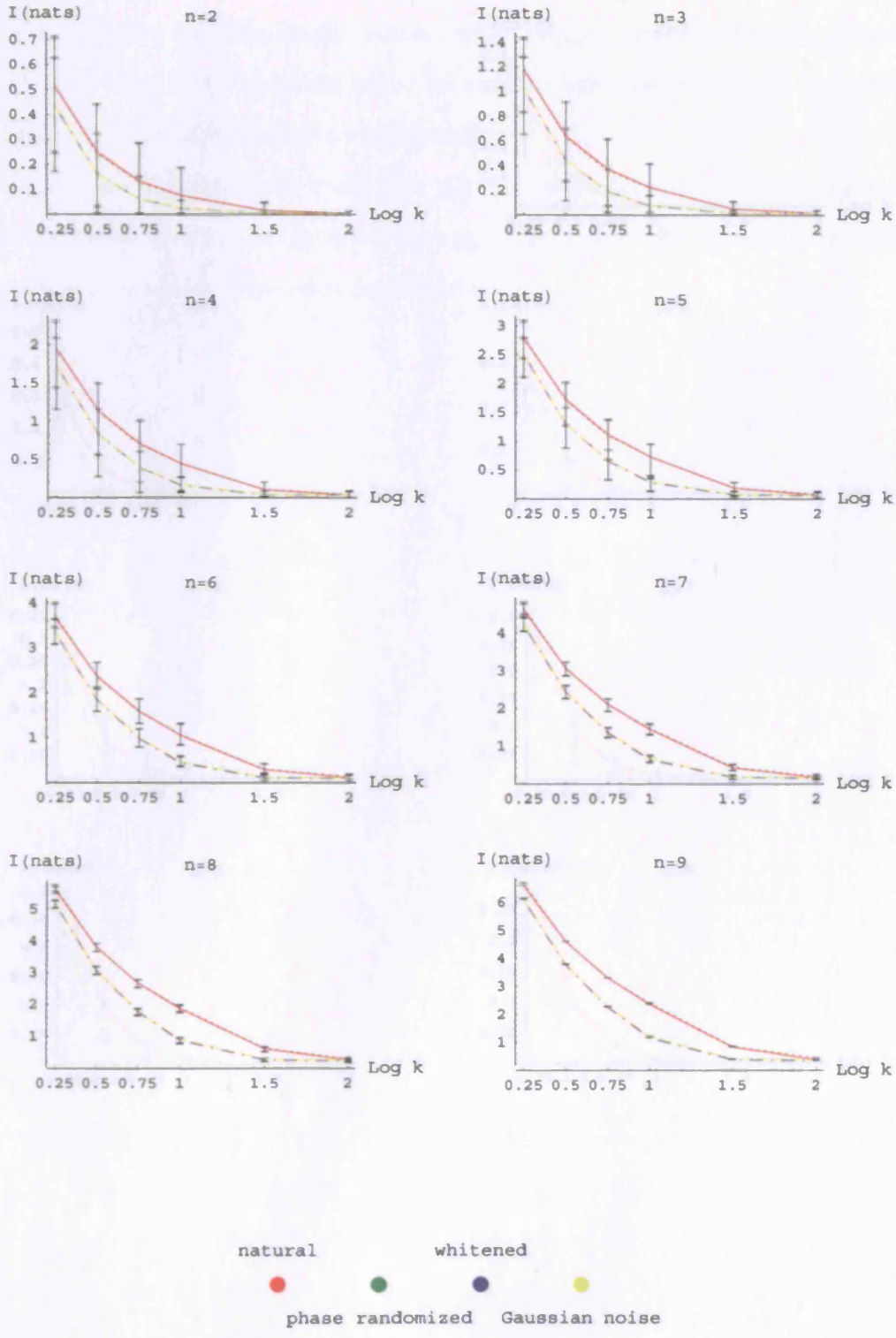


Figure 8.4: Illustrates how the shape index n-point dependency varies with distance between points k in different n-point configurations. Vertical axis: dependency between points in terms of total mutual information I(nats); horizontal axis: distance between points in n-point configuration k. Each curve in each chart represents different image classes. The error bars denote the maximum and minimum information scores for that n-point configuration.

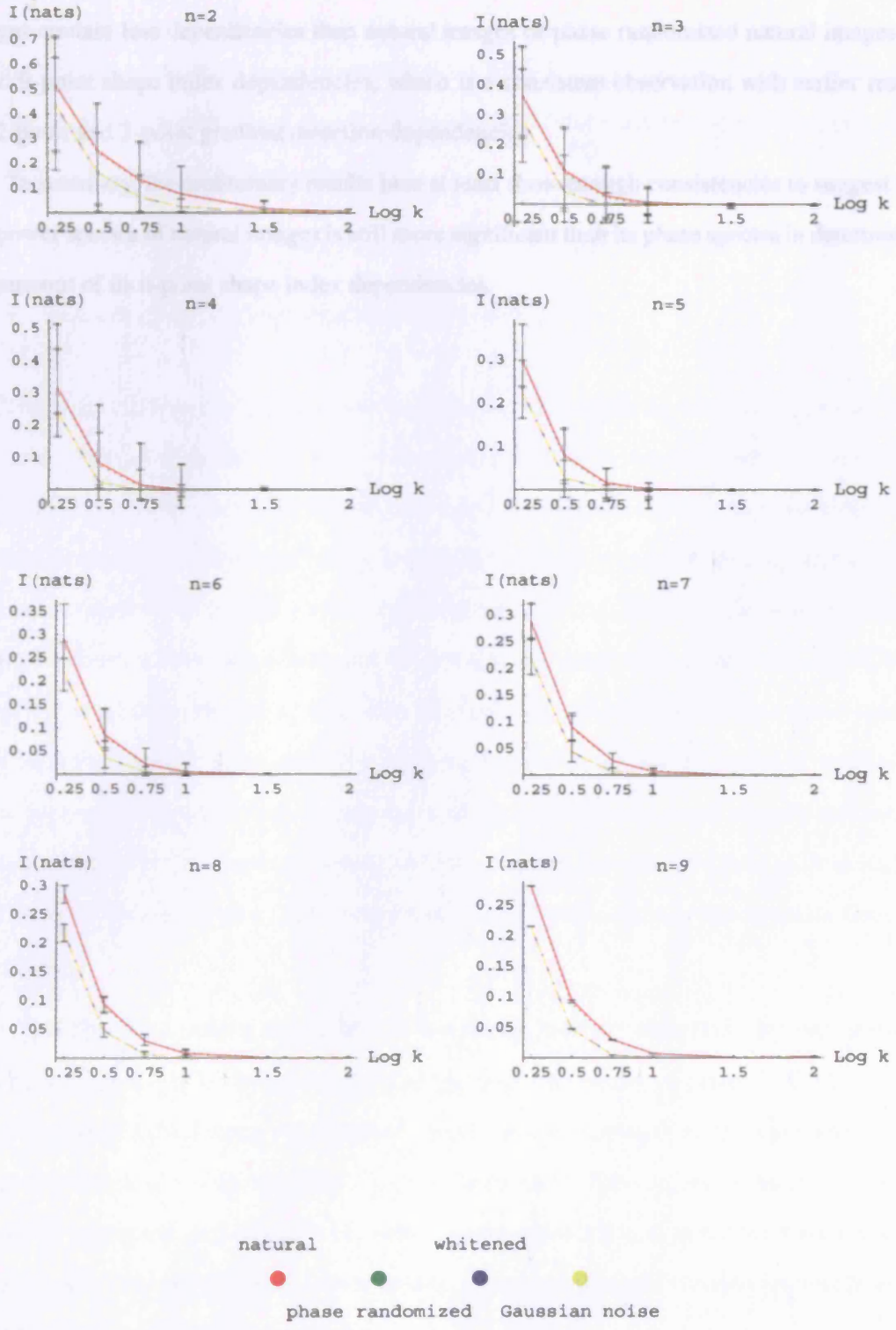


Figure 8.5: Illustrates how the shape index n-point interaction information varies with distance between points k in different n-point configurations. Vertical axis: interaction information I(nats); horizontal axis: distance between points in n-point configuration k. The error bars denote the maximum and minimum information scores for that n-point configuration.

Nevertheless, it is observed that for $k \leq 1.5$, whitened natural images and Gaussian noise images contain less dependencies than natural images or phase randomized natural images for up to 9-point shape index dependencies, which is a consistent observation with earlier results for 2-point and 3-point gradient direction dependencies.

In summary, the preliminary results here at least show enough consistencies to suggest that the power spectra of natural images is still more significant than its phase spectra in determining the amount of its n-point shape index dependencies.

8.2 Conclusions

In this section, summaries are presented for the motivation of this thesis, relevant work in the literature, and methodology. Lastly, a description of the main contributions of this thesis together with some final remarks are given.

8.2.1 Summary of Background and Motivation

In Chapter 2, introductory material was presented on the properties of light. Examples were provided demonstrating the way light interacts with objects in the environment; such a consistent relationship between light and objects in the environment enables the development of a reliable biological visual system. Next, the efficient coding hypothesis that the human visual system is optimised to encode natural visual stimuli was discussed with supporting evidence from the literature. The physiology and functions of the visual system were also discussed in Chapter 2 to motivate the use of first order Gaussian derivative operators to compute gradient direction dependencies in images. For example, studies of the human visual system reveal: i) the presence of neurons with specific orientation preferences to visual stimuli, and the existence of horizontal connections linking neurons with similar receptive field profiles such as similar orientation preference; and ii) the receptive field profiles of neurons resemble Gaussian derivatives.

To further motivate the main study of this thesis, previous vision theories on human visual perception, in particular Gestalt psychology, were presented in Chapter 2. The Gestaltists hypothesised that the human visual system uses a holistic approach to form perceptions and proposed several grouping principles. In this thesis, the holistic approach has motivated the search for synergetic dependencies between measurements made at multiple image locations. Recall that, in the case of 3-point measurements, synergy describes dependencies which are not attributable to any 2-point dependencies.

At the end of Chapter 2, computational techniques applied to vision were presented. For example, how features are extracted from images and the significance of first order derivative measurements in finding edges and object contours in images, i.e. exploring how the differential structure of images is related to image features.

8.2.2 Summary of Relevant Work

In Chapter 3, a number of important regularities in natural images that have been reported in the literature were discussed. For example, the scale-invariant property of natural images and how the average power spectrum of natural image ensembles follow a power law decay with respect to spatial frequency (averaged over orientation). Therefore, the significance of this regularity in the mean power spectrum of an ensemble of natural images was discussed in relation to the human visual system. For example, it has been reported that neurons in the retina exploit redundancies in natural visual stimuli. Furthermore, neurophysiological experiments have shown visual cortical neuron activity is suppressed when the visual system is being subjected to natural visual stimuli compared to random stimuli. Psychophysical experiments have also shown observers are better at discriminating synthetic images with the same power spectrum power law decay as natural image ensembles. Furthermore, computational experiments have derived filter functions based on efficient coding constraints: filters which reduce redundancy, maximise information transmissions, and increase independency between filters begin to resemble cortical visual receptive fields. Lastly in Chapter 3, geometric regularities found between edge segments—in relation to their relative edge orientations—in natural images were discussed.

8.2.3 Summary of Methodology

In Chapter 4, it was explained how, in order to evaluate the amount of dependency between gradient measurements, information-theoretic methods such as higher order mutual information are used. To obtain an estimate of the gradient direction dependencies, joint probability distributions of gradient directions were formed from extracting triples of gradient directions in images for different separations between the triples. To estimate Shannon's measure of entropy, entropy estimators were applied to the joint probability distributions. Furthermore, each entropy estimator has a particular variance, and therefore a bootstrap procedure was employed to assess such uncertainties.

In Chapter 5, it was shown how the four image ensembles for the main results were generated. The four main image classes studied were natural, phase randomized natural (equivalent individual power spectrum but random phase spectrum), whitened natural images (equivalent individual phase spectrum but power spectrum set to unity, i.e. flat), and Gaussian noise images (luminance function drawn from a Normal distribution). Triples of gradient directions, arranged in an equilateral triangular and a line configuration, were collected from each image

in each image class ensemble for different separations between measurements. The average dependencies between gradient directions, gradient magnitudes and strong gradient directions (i.e. at high gradient magnitude locations) were computed for an ensemble of images from each class and for individual images. Finally, estimates using the bootstrap procedure were made to compute the uncertainty in the information-theoretic calculations of the dependencies.

8.2.4 Main Findings

In Chapter 6, results for the main experiments on the four image classes: natural, phase randomized natural, whitened natural, and Gaussian noise are presented.

Section 6.1 shows that differences emerge between the four main image classes in the 1-point statistics of gradient directions. The histogram data of gradient directions obtained from the four image ensembles shows that natural images have a slight excess of vertically and horizontally oriented gradients; this is different to phase randomized natural images, which have a prevalence of vertically oriented gradients only. It has been argued that this bias is due to natural images being composed of sky (lighter) and ground (darker) regions. This contrast between the upper and lower regions results in vertical upward gradients. The histogram distribution of gradient directions in whitened images shows a slight prevalence of horizontally oriented gradients, and in Gaussian noise images there are no gradient orientation preferences (flat distribution). However, the 1-point statistics do not reveal anything about the dependencies between gradient directions at separate image locations, which is described by 2-point and 3-point statistics.

In Sections 6.2 and 6.4, results show that the average dependencies between pairs of (2-point) gradient directions, and between triples of (3-point) gradient directions are strongly dependent upon the mean power spectrum of the ensemble of natural images. This observation is based on the amount of 2-point and 3-point gradient direction dependency being the same in natural images and phase randomized natural images, as well as in whitened natural images and Gaussian noise images.

This implies that the power spectra of natural images are more significant than the phase spectra in contributing to gradient direction dependencies. At first this is a surprising result because the phase information of natural images has been shown to contribute to localised features such as lines and edges in natural images (Sec. 3.5.2). However, for images rich in textures and shading or with strong geometric forms, the power spectrum can contribute significantly to the appearance of an image (Sec. 3.3.4). It is therefore not necessary for an ensemble of images to

retain line and edge structure in order to keep, on average, the same amounts of gradient direction dependency. This finding also highlights the fact that gradient direction dependencies have been studied, and not dependencies between the orientation of edge segments, which have been studied by other authors [8, 9, 180]; although in similarity to [8] it is found that dependency between 2-point measurements falls rapidly with distance, especially beyond the effective size of the kernel operators.

Although a significant relationship is found between the gradient direction dependencies and the mean power spectrum of an ensemble of natural images, if instead the dependencies of i) gradient directions at strong gradient magnitude locations, and ii) gradient magnitudes are computed, the relationship fails and all four image ensembles: natural, phase randomized natural, whitened natural, and Gaussian noise have different amounts of dependency. Furthermore, the amount of 2-point and 3-point gradient direction dependency of individual natural images are different to their phase randomized versions. Therefore, the relationship found between the amount of gradient direction dependencies and the mean power spectrum of an ensemble of natural images cannot be extended to its strong gradient direction dependencies, its gradient magnitude dependencies, or individual natural images.

In Chapter 7 additional experiments are performed on different ensembles of images. In Section7.1.1, ensembles of images with varying power law mean power spectra (with respect to spatial frequency averaged over orientations) and random phases were examined. It was found that an approximate lawful relationship between the power law exponent of the ensemble's mean power spectrum and the spatial decay rate in the dependency between nearby gradient directions.

In Section 7.1.2, the average dependencies between two and three gradient directions of an ensemble of natural images is shown to be approximately the same as those found in an ensemble of dead leaves model images (generated from a collage of disks with a cubic power law decay of sizes and luminance values drawn randomly from an ensemble of natural images). It was also demonstrated that the relationship between the mean power spectrum and gradient direction dependencies does not exist for ensembles of images with scale-variant gradient direction dependencies. For example, in Section 7.1.3, an ensemble of images generated from a collage of constant-sized disks and random luminance values showed more gradient direction dependencies compared to its phase randomized version. The constant-sized disk image class

is an adaption of the dead leaves model which does not maintain the scale-invariant properties of natural images.

In a separate investigation to the main experiments, in Section 7.2, the 2-point gradient direction dependencies of natural images and phase randomized images as functions of both their spatial and angular separation highlighted extremely small differences between the two image classes. The results for natural images were consistent with the contents of natural images and studies by other authors of edge co-occurrence statistics [8, 9, 180]. For example, for an ensemble of natural images, it was found that the amount of dependency between a known gradient direction and a second gradient direction extends further if the observation is made perpendicular to the original gradient direction; this is consistent with an ensemble of natural images containing numerous extended edges. This does not make the relationship found between the average gradient direction dependencies and the mean power spectra of natural images incorrect, but rather it provides a constraint on when this relationship applies.

The 3-point gradient direction statistics in Section 6.4 for an equilateral triangular spatial configuration revealed that synergetic dependencies found in an ensemble of natural images are equivalent to its phase randomized version. Furthermore, the amount of synergy found in natural images was not large compared to the amount of redundancy, which suggests 3-point gradient direction interactions do not dominate in natural images. Moreover, in natural images, the amount of synergetic dependencies in gradient directions is comparable to that found for Gaussian noise images and whitened natural images although the distance between measurements at which the peak synergetic dependencies occur is greater in natural images than it is for Gaussian noise images.

In Section 7.3.3 investigations into the causes of synergy observed in the triangular configuration of triples of gradient directions in the four main image classes show that it is likely to be a consequence of the fact that the gradient of a scalar field has zero curl (zero curl also implies zero circulation around closed curves via Green's thoerem). This is based on experiments where curl free gradient vector fields are transformed into non-zero curl and non-zero divergence fields. For this latter type of gradient vector field, no synergetic 3-point gradient direction dependencies are found, and overall the amount of redundancy is increased.

Furthermore, in Section 7.3.1, it was found that for a line configuration of gradient directions, only whitened natural images and Gaussian noise images display synergetic dependencies

(natural and phase randomized natural images did not). These synergetic 3-point dependencies in a line configuration did not vanish upon adding curl to the gradient vector field because the gradient directions are not constrained along an open curve such as a line, unlike around a closed curve such as in the equilateral triangular configuration.

Lastly, in Section 8.1, preliminary work to establish the dependencies in shape index (second order derivative) up to 9-point configurations showed that synergetic dependencies are only found between shape index measurements for 3-point, 4-point, and 5-point configurations in all four image classes: natural, phase randomized natural, whitened natural, and Gaussian noise. This is consistent with observing 3-point synergy in gradient direction dependencies. Preliminary results also indicate that the mean power spectrum of an ensemble of natural images is generally more significant than its phase spectrum for nearby measurements. However, for 9-point configurations at larger distances (where the dependency is already relatively small), small differences emerge between all four image classes. But, further investigations are required to explain these differences in order to confirm that this is not an artifact of the experimental procedure.

8.2.5 Limitations of Work

In this section, the limitations of the work in this thesis are presented.

Information-theoretic Approach

The analysis in this thesis has emphasised the information-theoretic methods to evaluate the dependency between gradient direction measurements in images. Although this methodology suited the purpose of this thesis; namely the search for quantitative evidence that there exists an advantage in making multiple measurements in an image, there is a limitation to this type of evaluation.

The limitation refers to the difficulty in being able to describe how the interactions between gradient direction measurements relate to image features. One way to access further information is to examine the histogram data (from which the entropies and information measures are calculated). Once the entropy of, for example, the joint distribution between two gradient measurements is calculated, knowledge of the geometric configuration of the two gradient directions is lost unless careful consideration is paid to the method of analysis. This is because the entropy of a distribution does not necessarily indicate the shape of the distribution. Nevertheless, it is

possible to use the histogram data in conjunction with the information-theoretic calculations, as described in Section 7.4, and to perform methods of analysis that are more readily related to features in images, as performed in Section 7.2.

Overlap of Kernels

The results for different image classes show that the statistical dependencies found between gradient direction measurements occur when there is overlap between the derivative Gaussian kernels which are implemented to measure the gradient direction.

The results in this thesis quantify the dependence between gradient measurements in images, therefore it is necessary to make physical observations using non-zero sized apertures. Thus, the overlap between kernels of different measurements is an unavoidable consequence when measuring the gradient field of an image as described in Section 4.4.1 of the thesis, and is an intrinsic part of the dependence between physical measurements.

An alternative approach to the one adopted in this thesis is to blur the image first and then approximate derivatives by comparing neighbouring pixels (even in this instance there will still be overlap between measurements). However, this would lead to a modification of the data before differentiation is applied, which is undesirable and is not equivalent to the derivative of a sampled image. If it was possible to blur the image and then apply an infinitesimal differential operator (equivalent to a derivative Gaussian kernel with $\sigma = 0$) to this blurred image, then this would be equivalent to the method adopted in this thesis as described by Equation 4.23. In this latter case, it is apparent that we are not dealing with overlap between kernels, but the fact that differentiation and blurring of the observed image can not be separated, however, blurring is not an artifact because it is not possible to perform measurements at infinite resolution.

The reasons why taking the derivative of an image is not straightforward is because the differential operator is an ill-posed functional, therefore in this thesis linear scale-space theory [79, 165, 166] is followed whereby the operator (kernel) is regularised rather than the operand (image). Physical considerations of the behaviour of such measurements then leads to the use of a Gaussian kernel and its partial derivatives as being the appropriate family of kernels.

In other words, it is necessary to take the derivative of an 'observed' image, but such an 'observed' image is then not differentiable in the sense of Hadamard [164]. The solution to this problem, as proposed by Schwartz [163], is to regularise the data by convolution with a smooth

kernel, which is equivalent to convolving with the derivative of the kernel. Interestingly, from a physiological perspective, the kernel is equivalent to the role of a receptive field, and the similarity of Gaussian derivatives with the receptive field profiles of cells in the primate visual cortex were pointed out by Young [168, 169, 170, 171].

Negative interaction information

The quantity $I(X;Y;Z)$, described in Equation 4.11, can be negative, which is initially surprising because mutual information between two random variables is always greater than or equal to zero. However, this is not unexpected when we consider the meaning of $I(X;Y;Z)$.

$I(X;Y;Z)$ represents the reduction in mutual information between X and Y given knowledge of Z. If the mutual information between X and Y decreases with knowledge of Z, then $I(X;Y;Z)$ remains positive because this measures the reduction in the mutual information between X and Y when knowing Z. If the mutual information between X and Y increases with knowledge of Z, then we have that there is not a reduction in mutual information between X and Y with knowledge of Z, but the opposite, which is signified by $I(X;Y;Z) < 0$, which is the definition of synergy. Thus, if $I(X;Y;Z)$ measures the decrease in mutual information between X and Y with knowledge of Z, then when $I(X;Y;Z) < 0$, it implies there is an increase in mutual information between X and Y with knowledge of Z.

A simple example of a system displaying a synergetic dependency is one where three random variables A, B and C that can each take on values of 0 or 1, are related by $A = B + C \pmod 2$, which is better known as the logical operation: exclusive disjunction (EOR). In this system, it is clear that A is independent of B unless C has been determined, therefore we must have that $I(A;B) = 0$. Given the relationship imposed by EOR, we also have that $H(A|C) = 1$ bit (knowing only C leads to no less uncertainty about A), but in knowing B together with C, we can completely determine A, hence there is no uncertainty, so that $H(A|B,C) = 0$. Thus, using Equation 4.13, we can determine that $I(A;B|C) = 1$ bit, and therefore $I(A;B;C) = -1$ bit.

8.2.6 Final Remarks

The main findings in this thesis are that for an ensemble of natural images the average amount of their 2-point and 3-point gradient direction dependencies are determined by their mean power spectrum. This has also been shown for ensembles of noise (random phases) images with vary-

ing power law power spectra. Moreover, this relationship has been shown to not extend to the dependencies of gradient directions at high gradient magnitude locations, gradient magnitudes, individual natural images, or an ensemble images with scale-variant gradient direction dependencies.

Furthermore, no significantly greater amount of 3-point synergetic[1] gradient direction dependencies are found for natural images over other image classes, for example, its phase randomized and whitened versions, and Gaussian noise images. Therefore, these results do not provide quantitative evidence for a Gestalt holistic approach to vision. Although this does not remove the possibility that these weak synergetic dependencies could be part of a larger network of dependencies involving other measurements (e.g. shape index and gradient directions which both display 3-point synergetic dependencies). Thus, a particular state could contain a large number of weak signals, for example, from the dependencies of several different derivative order measurements, which could then result in a strongly dependent overall state. Furthermore, the lack of large synergetic dependencies between gradient directions does not necessarily mean that the information-theoretic framework presented—to evaluate quantitatively the advantages of multi-local analysis—would not yield stronger results for other measurements, which are not necessarily based on image derivatives.

Overall, this thesis has re-emphasised the importance of the statistical regularity found in the mean Fourier power spectrum of natural image ensembles, not for, as previously reported by other authors, correlations between image luminance values, but for the average dependencies between gradient directions—a first order derivative measurement. Initially, this is a surprising result because the power spectrum is said to contain only second order statistics. It is also surprising given the appearance of phase randomized natural images, which contain a lack of features such as edges and lines. However, the power spectrum can contain information about shading and texture patterns of an image. In this respect, it is not unreasonable that the average gradient direction dependencies are the same over an ensemble of natural images and phase randomized natural images. Finally, given that the relationship between the mean power spectrum and gradient direction dependencies applies to ensembles of natural images and not individual images, it is therefore a statistical relationship that is unlikely to be due to some mathematical correspondence to the Fourier power spectrum. This is in contrast to

[1]Recall that synergetic dependencies imply that knowledge of an additional measurement increases the mutual information between previous subsets of measurements.

the autocorrelation function, which has a direct mathematical correspondence to the power spectrum via a Fourier transform (Wiener-Khinchin theorem).

Appendix A

Properties of Light

A.1 Depth Cues

A change in luminance (edge) provides another cue to depth. The most common is partial occlusion. When an object partially occludes another, the occluding object is perceived to be closer to the observer than the occluded object, as shown in Figure A.1. This cue provides no information of how far one object is behind the other. When there is ambiguity as to which object is the occluder and which is the occluded, the amount of blurring of the edge boundary is an additional cue that can help to distinguish which object is nearer; this is known as occlusion edge blur [181].

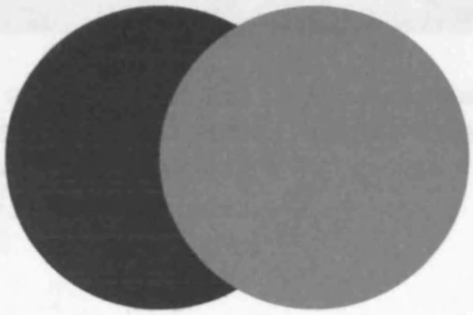


Figure A.1: A partially occluded object (black circle) appears further away from the observer than the occluding object (grey circle).

Texture gradients [34] can also be a cue to depth, an example of which is shown in Figure A.2. Surfaces have texture elements and by using the assumption that the elements are all the same size and shape [182] then gradually smaller projections of these elements can give the perception of an increase in depth as well as changes in surface orientation.

Figure A.2: Illustrates how texture elements of a brick wall diminish in size giving the illusion of depth even in a complicated scene such as this. Note also the presence of other depth cues such as changes in shape and vanishing points.

Appendix B

Human Visual System

B.1 Neuron Signals

Light reaching the photoreceptor cells must be converted into a neural signal (transduction). In fact, the percentage of signals successfully passed on once a photon is absorbed is around 70% [183]. Neurons within the brain send signals via biochemical reactions to other neurons. A typical neuron (see Figure B.1) will contain dendrites which get excited more or less—within some continuous range—by the amount of certain chemical substances known as neurotransmitters. This works by creating differences in the potential inside and outside the dendrite. The chemicals involved in creating this potential in humans tend to be positively charged sodium ions and potassium ions, and negatively charged chloride ions and protein molecules.

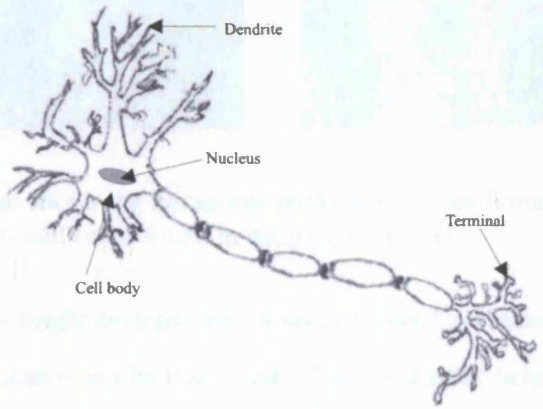


Figure B.1: Illustration of a cortical neuron.

At rest, i.e. when not sending a signal, a neuron will actually have a resting potential of -70 millivolts (mV). This negative potential is caused by an excess of sodium ions and potassium ions outside the neuron compared to inside the neuron (the set up of this gradient potential re-

quires energy which is provided by the photoreceptor cells when they are stimulated by light). The messages from all the dendrites arrive at the cell body which is like the 'call-centre' of the neuron. The cell body, by using its outer semi-permeable membrane, which allows potassium ions to pass through easily unlike sodium or chloride ions, integrates messages from the dendrites. This integrated signal is relayed along the neuron's axon using a series of action potentials known as spikes. The number of spikes in a given time interval is known as the neuron's firing rate and is a measure of the strength of the signal being transmitted by the neuron. Once the electric signal passes along the neuron's axon, it reaches the terminals where it is converted back to a chemical signal. At this point, more neurotransmitters are released across a synaptic gap to stimulate more dendrites from other neurons.

B.2 Photoreceptor Cells

Photoreceptor cells—cones and rods—are located at the back of the retina, as shown in Figure 2.6. Interestingly, the number of rods outweighs the number of cones considerably. There are over one hundred million rods compared to just six and half million cones [184] although the diameter of a cone is approximately three times that of a rod.

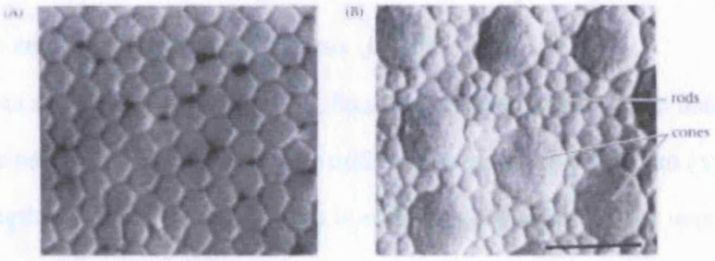


Figure B.2: Left image shows the hexagonal packing of cones found in the fovea, and right shows a mixture of rods and cones found in the periphery [185].

Cones are used for bright daylight conditions and colour, moreover cones can detect faster changes in light levels than rods which are used in low-level light. In humans, most of the cones are packed tightly into a region called the fovea which provides for the highest resolution in the centre of the visual field (see Fig. B.3).

The reason the photoreceptor cells appear at the back of the eye is because this provides contact with a row of dark cells (pigment epithelium) which replenishes vital molecules in the photoreceptor cells after being exposed to light, furthermore, this layer absorbs any stray light from being reabsorbed. However, the positioning of the photoreceptor cells means that light

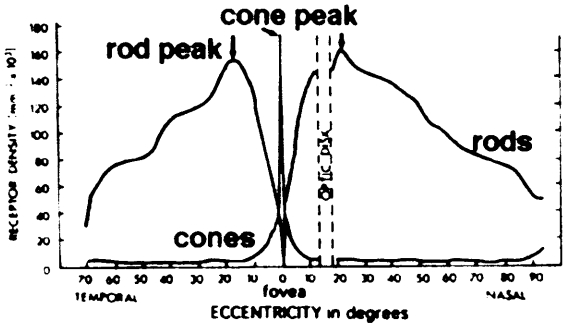


Figure B.3: Illustrates how the density of rods and cones in the human retina changes with eccentricity across the visual field [184].

must first pass through several layers of ganglion cells, amacrine cells, bipolar cells and horizontal cells before reaching any rods or cones (Fig. 2.6). In fact, much of the light entering the eye—around 90%—does not get absorbed by the photoreceptor cells, instead it is absorbed by non-receptive biological tissue (e.g. blood vessels) or simply passes through the photoreceptor cells without being absorbed [186].

B.3 Optic Chiasm and Lateral Geniculate Nucleus

The region of the human visual system between the retina and primary visual area (V1), contains the optic chiasm and lateral geniculate nucleus (LGN).

The axons of the retinal ganglion cells, bundle and form fibres which make up the beginning of the optic nerve. In moving along the optic nerve to the optic chiasm (χ) approximately a third of the length of the brain is covered. It is at the optic chiasm that the neural fibres leading from the right side of the left eye crossover to right side of the brain; fibers from the left side remain on the left side of the brain. This crossover of fibers is illustrated in Figure B.4.

Beyond the optic chiasm, is the optic tract and the lateral geniculate nucleus (LGN) which is part of the thalamus. The thalamus is the perceptual hub of the brain receiving information from other senses such as touch and sound. The LGN itself is a body of neurons arranged in several layers.

The receptive fields of LGN cells are rather similar to those of ganglion cells: roughly circular and have a centre-surround organisation[188]. However, unlike ganglion cells, which only change spatially, the LGN receptive fields change spatially and temporally. The behaviour of LGN cells can be modelled by the product of a Laplacian of a Gaussian in the spatial domain, and a first order derivative Gaussian in the temporal domain.

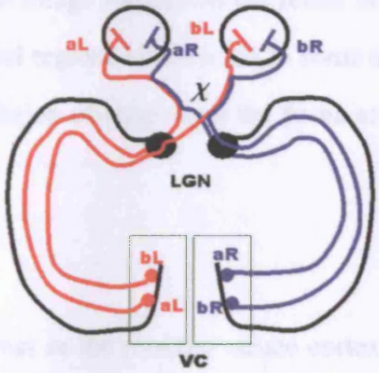


Figure B.4: illustrates the path light takes from the retina (aL,aR,bL,bR) to the striate cortex (VC). **a** denotes the left eye; **b** denotes the right eye. **L** denotes the *left* part of the retina which receives optical information from the *right* part of the visual field. **R** denotes the *right* part of the retina which receives information from the *left* part of the visual field. At χ, half the the nerve fibres crossover, so that the left LGN receives input from **aR** and **bR**. Consequently, the right LGN receives input from **aL** and **bL**. Eventually, optic information from the *left* visual field (**aL** and **bL**) arrives in the visual cortex(**VC**) in the *right* hemisphere of the brain. Thus optic information from the *right* visual field (**aR** and **bR**) arrives in the visual cortex(**VC**) in the *left* hemisphere of the brain [187].

The LGN is connects to V1 [77]. It is known that the lateral geniculate cells also receive feedback information from the cerebral cortex and input from the brainstem reticular formation (part of the brain associated with attention or arousal). Hypotheses suggest that this feedback loop adapts the receptive fields of LGN cells by using information processed in the cortex which is then relayed back to the LGN [189, 190].

The path leading from the eye to the lateral geniculate cells and then V1 is topographically organised. Thus areas which were originally next to each other in the eye (retina) will also be next to each other in the LGN. This is supported by the fact that if damage were sustained to a part of the human visual cortex then randomly delocalised blindness would not result. However, the central part of the image formed on the retina undergoes cortical magnification in comparison to more peripheral regions which leads to some distortion in the mapping [191]. This is due to the denser population of cones near the fovea and the low ganglion to receptor ratio.

B.4 Hypercolumns

In [192] it has been reported that in the monkey striate cortex, cells have a preference for a particular eye. In some layers one eye will dominant almost exclusively, e.g. in layer 4 (see Fig. 2.11). Therefore, cells in layer 4 are referred to as monocular. However, in layers above and below layer 4, there is a more equal share of cells between the eyes; these cells are often called binocular.

Within a hypercolumn there are two inner-columns that represent cells which are subject to Left or Right eye dominance. Within an individual hypercolumn, along one axis, the variation of orientation preference in cells is smooth [193], as shown in Figure B.5. The top surface shows the surface of V1 and below it, deeper levels of the cortex.

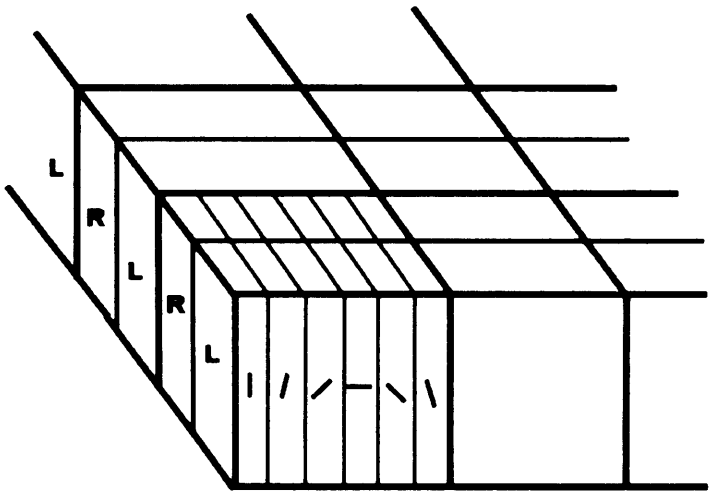


Figure B.5: Illustrates the hypercolumns of V1. Orientation preference rotates smoothly along the axis horizontally to the right. The top layer is the surface of the cortex, and along the other axis, cells which are subject to Left (L) or Right(R) eye dominance alternate. The dominance spreads horizontally in the direction of changing OP [194].

Appendix C

Human Visual Perception

C.1 Philosophical Implications of Gestalt Theory

A criticism of the Gestalt theory of perception is that it falls under the category of indirect realism (there were many advocates of this philosophy: Kant, Locke, Russell): the world is not perceived directly but through representations of objects although such representations are derived from real world data[1]. In other words, perceptions are caused by the intrinsic qualities of the environment—light and objects—that are interpreted to make inferences about the world. This appears to borrow from both unconscious inference and ecological optics. The reasons Gestalt theory is considered implausible is due to the problems of infinite regression and the need for a homunculus. Although these problems are inherent to indirect realism and not just Gestalt theory. Despite these objections, there have been attempts to continue the paradigm of internal representation by suggesting the world that is perceived is all in our head [195]. The theory suggests the boundary of this perceived world becomes the edge of our physical skull, beyond which, lies the real physical world. This would render the head that is known to be merely a tiny—perceptual—copy of our real one! Needless to say, such views have been met with fierce criticism.

A theory proposed by Köhler, known as isomorphism, posits that the brain produces electrical fields which are of the same shape as the projections of observed objects. But this theory implies that pictures are created in the head. However, who or what is observing this model in our head? Ultimately this leads to an infinite regress of beings making observations in the head [187]. Moreover, in [187], it explains that if the brain fields were really isomorphic, it would not be possible for perceptual grouping to occur since that would not be a true isomor-

[1] This is also known as representative realism.

phism of reality. However, it may not seem totally implausible that 3-D images could be created in the brain since studies using fMRI scans have shown V1 neurons 'lighting up' with patterns similar to the scene being observed, e.g. for a scene of a flickering checkerboard [55, 54]. However, the study used a simple pattern, is it possible for the brain to create complex natural scenes?

Interestingly, parallels can be drawn between Gestalt philosophy and another branch of research known as 'complexity', which deals with how complex systems evolve from very simple basic laws and interactions. In this paradigm, visual perception could be seen as a self organised critical system and when a critical point is reached a change in perception occurs. This provides a framework for multistability where it is a state between chaos and stability; such a transient condition is well described by self organised critical phenomenon. Visual perception could involve some form of complexity theory because there appears to be a discontinuity between the overall phenomenon (perception) and the individual elements (pixels) and their interactions (grouping rules). Indeed, it seems complexity arising from simplicity is analogous to the process of observing a visual scene and forming a perception. The perception, rather than being a simplification (in the sense of information theory) contains a complex meaning.

C.2 Holism

A holistic approach to perception contradicted the view of structuralists who hypothesised that perceptual understanding is built from small parts (atomism). Structuralists sought the elemental structures of mental items and believed that from associations between structures more complex groups are formed. However, Gestaltists suggested that wholes could not be split into such parts and their relations: "For an approach 'from above downward' i.e. from whole-properties downward towards subsidiary wholes and parts, individual parts ('elements') are not primary, not pieces to be combined in and-summations, but are parts of wholes" [61]. Therefore, experimental work in Gestalt theory tried to determine the nature of such wholes. For example, the visual percept in Figure C.1 bottom left illustrates a Necker cube. If the twelve lines of the cube are viewed in isolation, then all that is perceived are twelve lines orientated vertically or horizontally. However, if the same twelve lines are viewed together, positioned in a specific way, a perception of depth can be created. This phenomenon is emphasised by the circles surrounding the Necker cube (Fig. C.1 top with circles, and bottom left without circles). The

cube's position—inside of the four circles—produces the perception that the four circles are located at different depths. However, without the cube, the four circles appear located on a flat 2-D surface (Fig. C.1 bottom right). Another more complicated example of holistic perception, which involves the influence of memory, is presented in Section C.3.

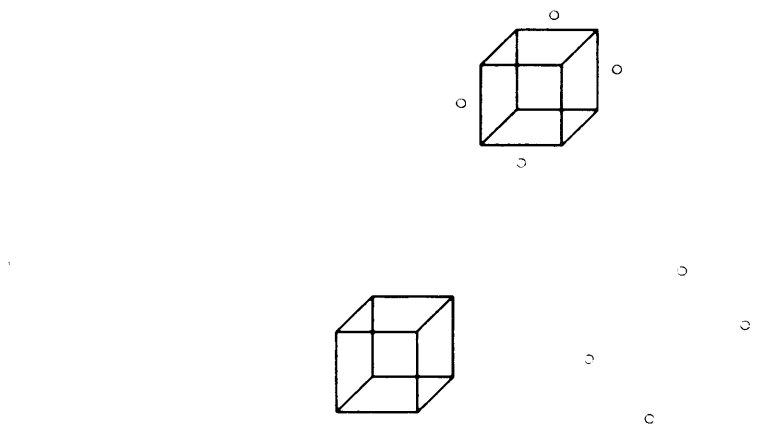


Figure C.1: The Necker cube (bottom left) illustrates how elements in a visual scene in isolation do not give rise to the same perception when grouped together. Although made up of only lines and circles the Necker cube provides percepts of depth for the surrounding circles (top). Remove the Necker cube (bottom right) and the perception of depth is lost and the circles appear to be on a 2-D flat surface.

C.3 Memory and Holistic Perception

When viewing Figure C.2 for the first time, the percept is of a random set of black splodges. Eventually, however, the scene of a dalmatian dog sniffing the ground is perceived, and every time thereafter it will be perceived almost immediately. Analysis of such figures and their eventual perceptions address important questions as to the role of memory in perception. The Gestalt proposal that perception works only from a 'top-down fashion' has been both supported and criticised. An argument used against 'top-down' processing occurring first is that if it were first the dalmatian in Figure C.2 should appear to us immediately. However, it is only a little later that the 'whole' is perceived. Alternatively, as suggested by Rock, it could be

that one searches the memory first and then takes into account the visual data; therefore, a 'top-down' process rather a 'bottom-up' process is occurring. However, to almost everyone, unconnected elements are first perceived. It is only a little later that the 'whole' is perceived. Explicitly, Rock states, 'The recourse to the memory is only done if the stimulus is similar to past experiences,....The memory is then accessed and is woven into the final percept so as to *enrich* it' [196], whereby to enrich is to use past experience to alter the current perception. Rock implies that when no similar past experience exists the current perception remains unaltered but it does not mean that no recourse to memory has been attempted.

Another view is that both top-down and bottom-up processes are working together almost simultaneously. For example, Julesz proposed that, "In real-life situations, bottom-up and top-down processes are interwoven in intricate ways, and the slogan of Gestalt psychologists that 'the whole is more than the sum of its parts'—a negation of the structuralist view of science—is probably true." [197].

Figure C.2: On first viewing, it is not clear what the image is meant to represent, eventually though, a dalmatian dog sniffing the ground is perceived. Once the percept has been observed it appears committed to memory so that any later recourse to the figure prompts a more immediate recognition of the dalmatian dog.

C.4 Common fate and Closure

Common fate is a grouping cue which posits that objects moving together will be grouped together, all other things being equal, as illustrated in Figure C.3. Good continuation is illustrated in Figure C.4a). Instead of perceiving four separate lines which all happen to meet at a common

point, 'the centre', two straight lines crossing over at the 'centre' are perceived. Thus, if all other factors are equal, elements that can be continued, will be grouped together. If there is an absence of the proximity cue then the good continuation cue is stronger when performing contour completion tasks [15]. However, quantitative descriptions of the good continuation cue are limited. Closure is demonstrated in Figure C.4b—two separate whole ellipses are perceived, rather than any continuation of one curve from the left hand side onto the right hand side.

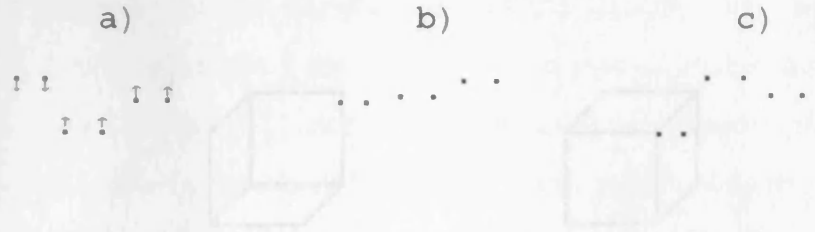


Figure C.3: Certain pairs of squares will be paired together, arrows point to the direction the squares will move in. Proximity is a conflicting factor in this stationary display, but would not be if actual motion could be achieved.

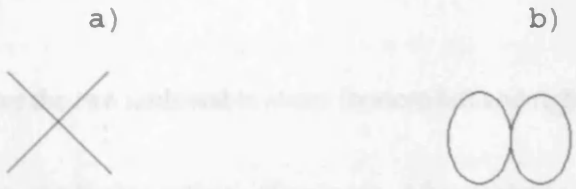


Figure C.4: Good continuation is demonstrated in a), rather than perceiving the ends of four straight lines all converging upon the same point, two lines crossing over is perceived. In b) the effect of good continuation is superseded by closure; two separate objects are perceived as adjoined.

C.5 Multistability

Multistability is present in the Necker cube (shown in Figure C.5 top); to make it more obvious the two alternative percepts are shown in Figure C.5 bottom left and right; the visual system has two competing hypotheses as to which face is the front face of the cube.

Köhler, a Gestalt psychologist tried to explain why multistability occurs. Köhler's theory—neural fatigue hypothesis—suggests that multistability is the result of neurons becoming 'tired' of firing after some time [198]. This hypothesis posits that each interpretation is a result of a pattern of neural activity. The perception experienced by the observer will be the

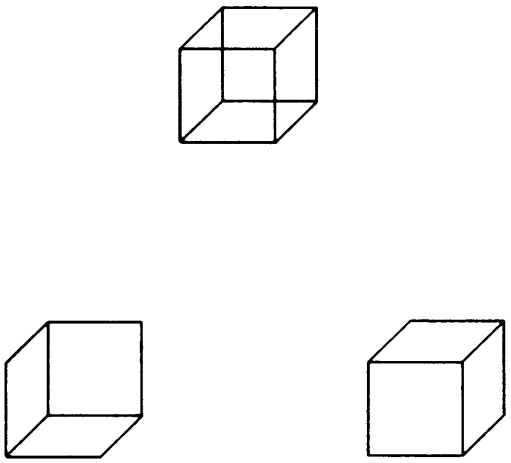


Figure C.5: Illustrates the two multistable states (bottom left and right) of the Neckercube (top).

interpretation that has the higher activity of neurons. After prolonged stimulation, neurons will experience a reduced firing rate as the neuron's stock of neurotransmitter substances (required to transport signals) reduces, which is neural fatigue. A model of neuron connections which could lead to multistable states when viewing a Necker cube is described next in Section C.6.

C.6 Description of a Multistable Neuron Network

In [199], a model of neuron connections which could account for multistability has been demonstrated for the perception of a Necker cube. The model essentially consists of two sets of neural activity S_1 and S_2 which are disjoint ($S_1 \bigcap S_2$). Each set represents the pattern of neurons responsible for a particular interpretation of the scene. If S_1 is activated more than S_2, it causes an inhibitory response in S_2 thereby reducing the firing rate of neurons in S_2, but in S_1 there is an excitatory response. Eventually neurons in S_1 dominate, however, after any prolonged period of excitation neural fatigue occurs, which leads to a reduction in the firing rate of neurons in S_1 and an increase in the firing rate of neurons in S_2. Eventually neurons in S_2 dominate, which leads to the observer experiencing the alternative scene interpretation. This cycle of excitatory

and inhibitory responses between S_1 and S_2 continues leading to changes in the observer's perception of the scene.

C.7 Competition between Gestalt Grouping Principles

How Gestalt grouping principles, described in Section 2.3.3, compete with each other was investigated by Hochberg and Hardy who found that up to a certain ratio, intensity cues would supersede proximity cues [200]. Tversky et al. suggested that the influence of closure is limited over good continuation and proximity [201]. Tversky et al. performed psychophysical experiments to test the ability of the HVS to perform contour grouping. The subjects were asked to identify closed contours which are edge fragments arranged into a circle, and open contours which are edge fragments arranged into an 'S' shape. These 'target' contours were hidden amongst other random edge fragments. The paper found that closure was not in itself important, and that good continuation and proximity were sufficient to account for both closed and open contours. However, the experimental results do not suggest that closure never influences perceptual organisation. Therefore, Tversky et al. conclude that, 'it is possible that closure mechanisms play an important role in perceptual organization' [201]. The results are not necessarily incompatible with Gestalt theory if contour grouping is regarded as a lower-level task compared with perceptual grouping.

The Gestalt principles presented are very simple examples and in each case the principles are viewed in isolation. For natural stimuli, the principles will be in direct competition with each other, making it much harder to determine the preference of the HVS for any one of principles.

Extensions to Gestalt Grouping Principles

The grouping principles of the Gestalts have been extended to include synchrony [202], common region [65], and connectedness [202]. The extended principle of connectedness states that elements that are within the same closed region of space will be grouped together. Moreover, elements which have connections to each other through additional elements are grouped together, rather than elements which are simply in closer proximity. Palmer suggests that proximity is in fact derived from connectedness and therefore connectedness is the underlying principle.

C.8 Quantifying Figural Goodness

Gestaltists stated that simplicity was the dominating factor for figural goodness. It was not possible to reduce figural goodness to quantitative properties such as the number of sides of a figure, or they simply could not find them.

Relatedly, other authors have tried to find quantitative measures. Properties such as the more sides a figure has the less 'good' it is. Differences of regularity are perceived by the HVS between figures with the same number of sides. For example, there is a perceived difference between a rectangle and a parallelogram; the square will be perceived as containing more 'goodness'. However, two groups: Attneave [1], and Hochberg and McAlister [203] independently proposed that figure 'goodness' could be quantified by considering the amount of information it would take to encode such figures. Both groups presented evidence that 'good' figures would require fewer bits to encode them than less 'good' figures. In the rectangle versus parallelogram example, all four interior angles of the rectangle are the same (90 degrees) in contrast to the parallelogram. Therefore, there are more regularities in the rectangle. Significantly, this difference can be expressed in terms of the amount of information needed to encode the rectangle, which is less than for the parallelogram.

However, such a description does not account for the anisotropy of perception. For example, perception is not invariant to rotations. This property was originally noted by Mach, who found that when viewing both a square, and another square rotated by 45 degrees, the latter is perceived—despite there being no geometrical difference between the two—as a diamond rather than a square rotated.

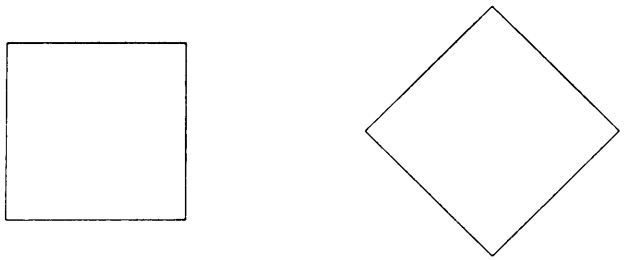


Figure C.6: Illustrates the effect of orientation upon perception. The square on the left is rotated by 45 degrees to give the figure on the right. The perception of the figure on the right is that of a diamond not a square, even though both figures are geometrically the same [196].

In [196], it was suggested that if an observer tilts their head by 45 degrees the original per-

ceptions of the square and diamond would remain, and therefore in his words, 'Only a change in perceived orientation affects the perceived shape' [196]. The Gestalts regarded this phenomenon as relational determination. Rather than consider the absolute geometrical properties of figures, the relations among properties and parts determine the perception. Note, in the example of the rotated square (diamond percept) in Figure C.6, our imagination could be used to rotate it again by 45 degrees to reform a perception of a square. Conversely, if the square is rotated by 45 degrees in our imagination alone, it would still appear as a square in our mind. This highlights the anisotropy in perception.

Appendix D

Computational Approach to Vision

D.1 Edge Detection Algorithms

A substantial amount of research has been dedicated to detecting edges and lines in images. Both features have been identified in the image processing community as important because the cause of an edge in an image is attributable to properties of surfaces of objects and the light field in the environment. The light reflected from a surface will change across different surface materials, orientations of those surfaces, and the amount of light falling onto those surfaces.

Computer algorithms performing edge detection utilise an edge operator which is applied across the whole image. The purpose of this edge operator is to have different responses depending on the relationship between neighbouring luminance values of pixels. For example, an edge operator should have a strong response when there are abrupt change between neighbouring luminance values; a weak response when there are gradual changes between neighbouring luminance values, and no response when luminance values between neighbouring pixels is uniform. Thus the overall output of an edge operator should be high at locations where there are sharply contrasting regions in the image, and low at all other locations. Examples of edge operators are shown in Figure D.1 and the application of one of the vertical edge operators to the luminance values of an image are shown in Figure D.2.

However, the example of an edge shown in Figure D.2(a) is rather idealised. Edges in images of natural scenes are not usually sharp step edges. There are several reasons which

(a)

-1	+1

(b)

+1
-1

(c)

-1	0	+1
-2	0	+2
-1	0	+1

(d)

+1	+2	+1
0	0	0
-1	-2	-1

Figure D.1: Left: (a) & (b) are examples of simple vertical and horizontal edge operators respectively, and right: (c) & (d) are Sobel edge operators.

(a)

1	1	1	1	1	1	1
1	1	1	1	1	1	1
5	5	5	5	1	1	1
5	5	5	5	1	1	1
5	5	5	5	1	1	1
5	5	5	5	1	1	1

(b)

0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	-4	0	0	0
0	0	0	-4	0	0	0
0	0	0	-4	0	0	0
0	0	0	-4	0	0	0

Figure D.2: Left: (a) is an example of an image displaying its luminance values as a numeric quantity. Right: (b) shows the values after applying (convolution) the vertical edge operator illustrated in Fig. D.1(a).

cause the smoothing of edges; it can be attributable to real effects in the environment or part of the measuring device (camera) which produces the image, for example, using a square aperture produces edges (artifacts) that are not present in the scene. Regardless of the cause, finding real edges rather than artifacts (noise), means that edge detection algorithms for natural images which adopt the strategy of applying edge operators at several scales should do a better job than one which use only one scale [87], an example of such an algorithm is described next.

Marr-Hildreth Zero-crossing Algorithm

The Marr-Hildreth zero-crossing algorithm [78] is a multi-scale edge finding algorithm that uses three spatial scales: fine, medium and coarse. The 'zero-crossing' refers to the use of second order differential operators to find edges; mathematically, this is equivalent to finding maxima in the first derivative of the image but its application, computationally, is different. At an edge, the second derivative of the intensity profile crosses zero at the location of the edge, and is flanked by positive and negative values on either side, depending on the direction of the edge. By computing zero-crossings rather than maxima, computation time can be saved because rather than having to pass four first-differential operators—horizontal, vertical, two diagonal—over a 2-D image, only one operator need be used, as illustrated in Figure D.3. Notice that the Marr-Hildreth edge operator, in Figure D.3(c), resembles the centre-surround organisation of receptive fields found in ganglion cells (see, Fig. 2.8) and LGN cells (both cells are circularly symmetric with a Mexican hat shape).

(a)

-1	+2	-1

(b)

-1
+2
-1

(c)

-1	-1	-1
-1	+8	-1
-1	-1	-1

Figure D.3: Vertical in (a), horizontal in (b) and two diagonal (not shown) edge operators can be replaced by the discrete version of the Marr-Hildreth edge operator in (c) which measures the second derivative of a 2-D image averaged over all directions.

Canny Algorithm

Despite its multi-scale approach, the Marr-Hildreth algorithm still suffers from finding artifacts or poor localisation of edges; an alternative algorithm was developed by Canny. The Canny edge detection algorithm [204] uses a set of eight differently oriented first order differential operators and detects edges as peaks in the output of these operators. This multi-directional operator scheme provides better detection of edges than the Marr-Hildreth algorithm and according to Canny is optimal at finding the most edges by reducing error rates, and localizing edges as closely as possible to the actual edge position.

The optimisation part of the algorithm finds local maxima by comparing pixels along the direction of the gradient and suppressing any non-maximal responses. This reduces the number of false edges. There is then a two level—high and low—threshold stage; pixel values with gradient magnitude higher than the high-threshold level are detected as edges, and if a pixel value lies on a contour it is also classed as being on an edge if its gradient magnitude value is between the high- and low-threshold values. This allows the algorithm to preserve weaker edges, but the setting of threshold values does affect the performance of the algorithm by reporting false edges or missing edges.

Scale-space Algorithm

Neither algorithm—Marr and Hildreth, or Canny—properly integrates the output of the edge operators at different scales. The problem is determining which edges are part of the same edge across the different scales. This is because the same edge may appear at different locations in the image depending on the scale. Therefore, a deficiency in both algorithms is accurately finding the location of an edge. To improve performance, it has been shown that edge operators should act at a continuum of scales spanning the size of structures in the image from fine scale to coarse, not just three scales as in the Marr-Hildreth algorithm. This scheme was proposed by Witkin (Scale-space filtering) in his scale-space algorithm [166] and was part of a larger concept known as the scale-space of an image, which is the stack of images at different scales described by Koenderink [79] and Lindeberg [87].

Bibliography

[1] F Attneave. Informational aspects of visual perception. *Psychological Review*, 61:183–193, 1954.

[2] HB Barlow. Possible principles underlying the transformation of sensory messages. In WA Rosenblith, editor, *Sensory communication*. MIT, Cambridge, Mass, 1961.

[3] HB Barlow. Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, 1:371–394, 1972.

[4] V Caselles, B Coll, and J Morel. Geometry and color in natural images. *Journal of Mathematical Imaging and Vision*, 16(2):89–105, 2002.

[5] E Switkes, MJ Mayer, and JA Sloan. Spatial frequency analysis of the visual environment: Anisotropy and the carpentered environment hypothesis. *Vision Research*, 18:1393–1399, 1978.

[6] RJ Baddeley and PJB Hancock. A statistical analysis of natural images matches psychophysically derived orientation tuning curves. *Proceedings: Biological Sciences*, 246:219–223, 1991.

[7] DM Coppola, HR Purves, AN McCoy, and D Purves. The distribution of oriented contours in the real world. *Proceedings of the National Academy of Sciences*, 95(7):4002–4006, 1998.

[8] JH van Hateren and A van der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of Royal Society London B.*, 265:359–366, 1998.

[9] M Sigman, GA Cecchi, CD Gilbert, and MO Magnasco. On a common circle: Natural scenes and gestalt rules. *Proceedings of the National Academy of Sciences*, 98(4):1935–1940, 2001.

[10] CC Chow, DZ Jin, and A Treves. Is the world full of circles? *Journal of Vision*, 2(8):571–576, 2002.

[11] CD Gilbert and TN Wiesel. Columnar specificity of intrinsic horizontal and cortico-cortical connections in cat visual cortex. *Journal of Neuroscience*, 9:2432–2442, 1989.

[12] KE Schmidt, R Goebel, S Lowel, and W Singer. The perceptual grouping criterion of colinearity is reflected by anisotropies of connections in the primary visual cortex. *European Journal of Neuroscience*, 9(5):1083–1089, 1997.

[13] R Malach, Y Amir, M Harel, and A Grinvald. Relationship between intrinsic connections and functional architecture revealed by optical imaging and in vivo targeted biocytin injections in primate striate cortex. *Proceedings of the National Academy of Sciences*, 90(22):10469–10473, 1993.

[14] F Orabona, G Metta, and G Sandini. Learning association fields from natural images. In *Proceedings of IEEE CVPRW*, page 172, 2006.

[15] DJ Field, A Hayes, and RF Hess. Contour integration by the human visual system: evidence for a local "association field". *Vision Research*, 33(2):173–193, 1993.

[16] C Zetzsche, E Barth, G Krieger, and B Wegmann. Neural network models and the visual cortex: The missing link between cortical orientation selectivity and the natural environment. *Neuroscience Letters*, 228(3):155–158, 1997.

[17] Y Petrov and L Zhaoping. Local correlations, information redundancy, and sufficient pixel depth in natural images. *Journal of Optics Society of America A*, 20(1):56–66, 2003.

[18] DJ Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of Optical Society of America A*, 4:2379–2394, 1987.

[19] JH van Hateren and DL Ruderman. Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proceedings of Royal Society of London B: Biological Sciences*, 265:2315–2320, 1998.

[20] MV Srinivasan, SB Laughlin, and A Dubs. Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society London B: Biological Sciences*, 216(1205):427–459, 1982.

[21] B Julesz. Visual pattern discrimination. *IRE Trans. on Information Theory*, IT-8:84–92, 1962.

[22] JR Bergen and EH Adelson. Visual texture segmentation based on energy measures. *Journal of the Optical Society of America A*, 3:99, 1986.

[23] DC Knill, DJ Field, and D Kersten. Human discrimination of fractal images. *Journal of Optics Society of America A*, 7:1113–1123, 1990.

[24] CA Parraga, T Troscianko, and DJ Tolhurst. The human visual system is optimised for processing the spatial information in natural visual images. *Current Biology*, 10:35–38, 2000.

[25] E Brunswik and J Kamiya. Ecological cue-validity of 'proximity' and of other gestalt factors. *American Journal of Psychology*, 66:20–32, 1953.

[26] JP Jones and LA Palmer. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6):1233–1258, 1987.

[27] D Kelly. Information capacity of a single retinal channel. *IEEE Transactions on Information Theory*, 8.

[28] BA Olshausen and DJ Field. Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, 14:481–487, 2004.

[29] D Attwell and SB Laughlin. An energy budget for signaling in the grey matter of the brain. *Journal of Cerebal Blood Flow & Metabolism*, 21:1133–1145, 2001.

[30] P Lennie. The cost of cortical computation. *Current Biology*, 13:493–497, 2003.

[31] WE Vinje and JL Gallant. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287:1273–1276, 2000.

[32] DS Reich, F Mechler, and JD Victor. Formal and attribute-specific information in primary visual cortex. *Journal of Neurophysiology*, 85:305–318, 2000.

[33] WE Vinje and JL Gallant. Natural Stimulation of the Nonclassical Receptive Field Increases Information Transmission Efficiency in V1. *Journal of Neuroscience*, 22(7):2904–2915, 2002.

[34] JJ Gibson. *Perception of the visual world.* Houghton Mifflin, Boston, 1950.

[35] JJ Gibson. *The ecological approach to visual perception.* Houghton Mifflin, Boston, 1979.

[36] D Marr. Representing visual information. *Lectures on Mathematics in the Life Sciences*, 10:101–180, 1978.

[37] D Marr. Visual information processing: The structure and creation of visual representations. *Proceedings of the Royal Society of London, B*, 290:199–218, 1980.

[38] D Marr. *Vision.* Freeman, San Francisco, 1982.

[39] D Marr and HK Nishihara. Theory of edge detection. *Proceedings of the Royal Society of London, B*, 200:269–294, 1978.

[40] JJ Koenderink, AJ van Doorn, and AML Kappers. Surface perception in pictures. *Psychophysics and Perception*, 52(5):487–496, 1992.

[41] JJ Koenderink, AJ van Doorn, and AML Kappers. Pictorial surface attitude and local depth comparisons. *Psychophysics and Perception*, 58(2):163–173, 1996.

[42] JJ Koenderink, AJ van Doorn, C Christou, and JS Lappin. Shape constancy in pictorial relief. In *ECCV '96: Proceedings of the International Workshop on Object Representation in Computer Vision II*, pages 151–164, London, UK, 1996. Springer-Verlag.

[43] JJ Koenderink, SC Pont, and AJ van Doorn. A new twist to the 'shading cue'. *Journal of Vision*, 6(6):260–260, 2006.

[44] SC Pont, AJ van Doorn, and JJ Koenderink. Light field matching. *Journal of Vision*, 4(8):122–122, 2004.

[45] WH Ittelson. Size as cue to distance. *American Journal of Psychology*, 64:54–67, 1950.

[46] W Epstein. The known-size apparent-distance hypothesis. *American Journal of Psychology*, 74:333–346, 1965.

[47] RP O'Shea, SG Blackburn, and H Ono. Contrast as a depth cue. *Vision Research*, 34(12):1595–1604, 1994.

[48] RP O'Shea, DG Govan, and R Sekuler. Blur and contrast as pictorial depth cues. *Perception*, 26:599–612, 1997.

[49] G Mather. Image blur as a pictorial depth cue. In *Proceedings: Biological Sciences*, volume 263, pages 169–172. The Royal Society, 1996.

[50] CT Morgan. *Physiological Psychology*. McGraw-Hill series in Psychology. Kogakusha Co, Ltd, Tokio, Japan, 3 edition, 1965.

[51] DH Hubel. *Eye, Brain, And Vision*. Scientific American Library, New York, 1987.

[52] SW Kuffler. Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, 16:37–68, 1953.

[53] CD Gilbert and TN Wiesel. Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *Neuroscience*, 9:2432–2442, 1989.

[54] Z Kourtzi and N Kanwisher. Cortical regions involved in perceiving object shape. *Neuroscience*, 20(9):3310–3318, 2000.

[55] S Murray. Shape perception reduces activity in human primary visual cortex. *PNAS*, 99(23):15164–15169, 2002.

[56] MK Kapadia, G Westheimer, and CD Gilbert. Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *Neurophysiology*, 84(4):2048–2062, 2000.

[57] MK Kapadia, M Ito, CD Gilbert, and G Westheimer. Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. *Neuron*, 15(4):843–856, 1995.

[58] JI Nelson and BJ Frost. Intracortical facilitation among co-oriented, co-axially aligned simple cells in cat striate cortex. *Brain Research*, 61(1):54–61, 1985.

[59] WH Bosking, Y Zhang, B Schofield, and D Fitzpatrick. Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *NeuroScience*, 17(6):2112–2127, 1997.

[60] S Thorpe, D Fize, and C Marlot. Speed of processing in the human visual system. *Nature*, 381:520–522, 1996.

[61] Wertheimer. *A sourcebook of Gestalt psychology*, chapter Laws of organization in perceptual forms, pages 71–88. London, UK: Routledge and Kegan Paul, 1938.

[62] RL Gregory. *Mind In Science*. Penguin Books, 1981.

[63] DC Knill, D Kersten, and A Yuille. *Perception as Bayesian Inference*. Cambridge University Press, 1996.

[64] K Koffka. *Principles of Gestalt psychology*. New York: Harcourt, Brace, & World, 1935.

[65] SE Palmer. Common region: A new principle of perceptual grouping. *Cognitive Psychology*, 24(3):436–447, 1992.

[66] Palmer S and I Rock. Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin and Review*, 1(1):29–55, 1994.

[67] S Wolfram. *Mind In Science*. Wolfram Media, Inc., 2003.

[68] M Kubovy and AO Holcombe. On the lawfulness of grouping by proximity. *Cognitive Psychology*, 35:71–98, 1998.

[69] HB Barlow. The efficiency of detecting changes of density in random dot patterns. *Vision Research*, 18:637–650, 1978.

[70] T Oyama. Perceptual grouping as a function of proximity. *Perceptual and Motor Skills*, 13:305–306, 1961.

[71] JH Elder and S Zucker. A measure of closure. *Vision Research*, 34:3361–3370, 1994.

[72] PU Tse. Volume completion. *Cognitive Psychology*, 39:37–68, 1999.

[73] J Shi and J Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[74] JH Elder and RM Goldberg. Ecological statistics of gestalt laws for the perceptual organization of contours. *Journal of Vision*, 2(4):324–353, 2002.

[75] B Julesz. Textons, the elements of texture perception and their interactions. *Nature*, 290:91–97, 1981.

[76] DH Hubel and TN Wiesel. Receptive fields of single neurons of single neurons in the cat's visual cortex. *Journal of Physiology*, 148:574–591, 1962.

[77] DH Hubel and TN Wiesel. Receptive fields, binocular interaction, functional architecture in the cat's visual cortex. *Journal of Physiology*, 160:106–154, 1962.

[78] D Marr and EC Hildreth. Theory of edge detection. *Proceedings of the Royal Society of London, B*, 207:187–217, 1980.

[79] JJ Koenderink. The structure of images. *Biological Cybernetics*, 50(5):363–370, 1984.

[80] L Kitchen and A Rosenfeld. Gray-level corner detection. *Pattern Recognition Letters*, 1(2):95–102, 1982.

[81] L Dreschler and HH Nagel. Volumetric model and 3D-trajectory of a moving car derived from monocular tv-frame sequences of a street scene. *Computer Vision Graphics and Image Processing*, 20(3):199–228, 1982.

[82] JA Noble. Finding corners. *Image and Vision Computing*, 6(2):121–128, 1988.

[83] JJ Koenderink and W Richards. Two-dimensional curvature operators. *Journal of the Optical Society of America*, 5(7):1136–1141, 1988.

[84] R Deriche and G Giraudon. Accurate corner detection: An analytical study. In *Proceedings 3rd International Conference on Computer Vision*, pages 66–70, Osaka, Japan, 1990.

[85] J Blom. *Topological and Geometrical Aspects of Image Structure.* PhD thesis, Department of Medical Physics, University Utrecht, NL-3508 Utrecht, Netherlands, 1992.

[86] LMJ Florack, BM ter Haar Romeny, JJ Koenderink, and MA Viergever. Scale and the differential structure of images. *Image and Vision Computing*, 10:376–388, 1992.

[87] T Lindeberg. *Scale-Space Theory in Computer Vision.* The Kluwer International Series in Engineering and Computer Science. Kluwer Academic Publishers, Dordrecht, Netherlands, 1994.

[88] H Wang and M Brady. Real-time corner detection algorithm for motion estimation. *Image and Vision Computing*, 13(9):695–703, 1995.

[89] T Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):77–116, 1998.

[90] Z Zheng, H Wang, and EK Teoh. Analysis of gray level corner detection. *Pattern Recognition Letters*, 20(14):149–162, 1999.

[91] CE Shannon and W Weaver. *The Mathematical Theory of Communication.* University of Illinois Press, 1949.

[92] MC Morrone, JR Ross, DC Burr, and RA Owens. Mach bands are phase dependent. *Nature*, 324:250–253, 1986.

[93] MC Morrone and RA Owens. Feature detection from local energy. *Pattern Recognition Letters*, 6(5):303–313, 1987.

[94] M Morrone and DC Burr. Feature detection in human vision: A phase-dependent energy model. *Royal Society of London Proceedings Series B*, 235:221–245, 1988.

[95] GS Heese and MA Georgeson. Edges and bars: where do people see features in 1-D images. *Vision Research*, 45:507–525, 2005.

[96] GJ Burton and IR Moorhead. Color and spatial structure in natural scenes. *Applied Optics*, 26:157–170, 1987.

[97] DJ Field. Scale-invariance and self-similar 'wavelet' transforms: an analysis of natural scenes and mammalian visual systems. In *Wavelets, Fractals, and Fourier Transforms*, pages 151–193. Clarendon Press, 1993.

[98] DL Ruderman and W Bialek. Statistics of natural images: Scaling in the woods. *Physics Review Letters*, 73:814–817, 1994.

[99] DW Dong and JJ Atick. Statistics of natural time-varying images. *Network Computation in Neural Systems*, 6(3):345–358, 1995.

[100] A van der Schaaf and JH van Hateren. Modeling the power spectra of natural images: statistics and information. *Vision Research*, 36:2759–2770.

[101] ER Kretzmer. Statistics of television signals. *Bell Systems Technology*, 31:751–763, 1952.

[102] NG Deriugin. The power spectrum and the correlation function of the television signal. *Telecommunications*, 1:1–12, 1956.

[103] DL Ruderman. Origins of scaling in natural images. *Vision Research*, 37(23):3385–98, 1997.

[104] AB Lee and D Mumford. An occlusion model generating scale-invariant images. In *Proceedings of IEEE Workshop on Statistical and Computational Theories of Vision*, 1999.

[105] RM Balboa, CW Tyler, and NM Grzywacz. Occlusions contribute to scaling in natural images. *Vision Research*, 41(7):955–64, 2001.

[106] W Hsiao and R Millane. Effects of occlusion, edges, and scaling on the power spectra of natural images. *Journal of Optics Society America*, A(22):1789–1797, 2005.

[107] DJ Tolhurst, Y Tadmor, and T Chao. Amplitude spectra of natural images. *Ophthalmic and Physiological Optics*, 12(2):229–232, 1992.

[108] MS Langer. Large-scale failures of f^{-a} scaling in natural image spectra. *Journal of Optics Society America A*, 17:28–33, 2000.

[109] Y Tadmor and DJ Tolhurst. Both the phase and the amplitude spectrum may determine the appearance of natural images. *Vision Research*, 33:141–145, 1993.

[110] CA Párraga, T Troscianko, and DJ Tolhurst. The human visual system is optimised for processing the spatial information in natural visual images. *Current Biology*, 10(1):35–38, 2000.

[111] I Juvells, S Vallmitjana, A Carnicer, and J Campos. The role of amplitude and phase of the Fourier transform in the digital image processing. *American Journal of Physics*, 59:744–748, 1991.

[112] JH van Hateren. Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation. *Journal of Comparative Physiology*, A(171):157–170, 1992.

[113] M Kardar and A Zee. Information optimization in coupled audio-visual cortical maps. *Proceedings of the National Academy of Sciences*, 99(25):15894–15897, 2002.

[114] JJ Atick and AN Redlich. What does the retina know about natural scenes? *Neural Computation*, 4(2):196–210, 1992.

[115] DJ Tolhurst and Y Tadmor. Band-limited contrast in natural images explains the detectability of changes in the amplitude spectra. *Vision Research*, 37(23):3203–3215, 1997.

[116] CA Parraga and DJ Tolhurst. The effect of contrast randomisation on the discrimination of changes in the slopes of the amplitude spectra of natural scenes. *Perception*, 29(9):1101–1116, 2000.

[117] RA Frazor and WS Geisler. The statistics of local contrast and mean luminance in natural images. *Journal of Vision*, 4(8):48–48, 2004.

[118] SB Laughlin. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch*, 36(C):910–12, 1981.

[119] G Sclar, P Lennie, and DD DePriest. Contrast adaptation in striate cortex of macaque. *Vision Research*, 29:747–755, 1989.

[120] DG Albrecht and WS Geisler. Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Visual Neuroscience*, 7:531–546, 1991.

[121] AB Bonds. Temporal dynamics of contrast gain in cells of the cat striate cortex. *Nature Vision Neuroscience*, 21(6):239–255, 1991.

[122] JL Gardner, P Sun, AR Waggoner, K Ueno, K Tanaka, and K Cheng. Contrast adaptation and representation in human early visual cortex. *Neuron*, 47:607–620, 2005.

[123] LP Kadanoff. Scaling laws for Ising models near T_c. *Physics*, 2:263–72, 1966.

[124] DL Ruderman. Statistics of natural images. *Network: Computation in Neural Systems*, 5:517–548, 1994.

[125] ET Jaynes. Information theory and statistical mechanics I. *Physics Review*, 106:620–630, 1957.

[126] E Schneidman, MJ Berry II, R Segev, and W Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440:1007–12, 2005.

[127] JG Daugman. Entropy reduction and decorrelation in visual coding by oriented neural receptive fields. *IEEE Transactions on Biomedical Engineering*, 36:107–114, 1989.

[128] R Baddeley, LF Abbott, MC Booth, F Sengpiel, T Freeman, EA Wakeman, and ET Rolls. Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proceedings of the Royal Society London B Biology Science*, 264:1775–1783, 1998.

[129] DJ Field. What the statistics of natural images tell us about visual encoding. *SPIE*, 1077:269–276, 1989.

[130] AJ Bell and TJ Sejnowski. The 'independent component' of natural scenes are edge filters. *Vision Research*, 37(23):3327–3338, 1997.

[131] BA Olshausen and DJ Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.

[132] MS Lewicki and BA Olshausen. A probabilistic framework for the adaptation and comparison of image codes. *Journal of Optics Society America A*, 16(7):1587–1601, 1999.

[133] EP Simoncelli and BA Olshausen. Natural image statistics and neural representation. *Annual Reviews of Neuroscience*, 24:1193–1215, 2001.

[134] MS Lewicki and BA Olshausen. A probabilistic framework for the adaptation and comparison of image codes. *Journal of the Optical Society of America*, 16:1587–1601, 1999.

[135] B Wegmann and C Zetzsche. Statistical dependence between orientation filter outputs used in a human-vision-based image code. In Murat Kunt, editor, *Proceedings of Visual*

Communication Image Processing, volume 1360, pages 909–922. Lausanne, Switzerland: SPIE, 1990.

[136] EP Simoncelli. Statistical models for images: Compression, restoration and synthesis. In *31st Asilomar Conference on Signals, Systems and Computers*, pages 673–678, Pacific Grove, CA, 1997. IEEE Computer Society.

[137] EP Simoncelli and O Schwartz. Image statistics and cortical normalization models. In *Advances in Neural Information Processing*, pages 153–159, Cambridge, Massachusetts, 1999. MIT Press.

[138] KD Miller. A model for the development of simple-cell receptive fields and the arrangement of orientation columns through activity dependent competition between on- and off-center inputs. *Journal of Neuroscience*, 14:409–441, 1994.

[139] MGA Thomson. Visual coding and the phase structure of natural scenes. *Network: Computation in Neural Systems*, 10(2):123–132, 1999.

[140] DJ Field. What is the goal of sensory coding? *Neural Computation*, 6(4):559–601, 1994.

[141] RJ Summers and MGA Thomson. Nonuniform phase perturbations in natural images. *Perception*, 28:Supplement 127, 1999.

[142] JD Victor and MM Conte. The role of high-order phase correlations in texture processing. *Vision Research*, 36(11):1615–1631, 1996.

[143] B Julesz and RA Schumer. Early visual perception. *Annual Review of Psychology*, 32:575–627, 1981.

[144] AV Oppenheim and JS Lim. The importance of phase in signals. *Proceedings of IEEE*, 69:529–541, 1981.

[145] MJ Morgan, J Ross, and A Hayes. The relative importance of local phase and local amplitude in patchwise image reconstruction. *Biological Cybernetics*, 65(2):113–119, 1991.

[146] LN Piotrowski and FW Campbell. A demonstration of the visual importance and flexibility of spatial frequency amplitude and phase. *Perception*, 11:337–446, 1982.

[147] MGA Thomson, DH Foster, and RJ Summers. Human sensitivity to phase perturbations in natural images: a statistical framework. *Perception*, 29:1057–1069, 2000.

[148] G Felsen, J Touryan, F Han, and Y Dan. Cortical sensitivity to visual features in natural scenes. *PLoS Biology*, 3(10), October 2005.

[149] JJ Knierim and DC van Essen. Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *Journal of Neurophysiology*, 67(4):961–980, 1992.

[150] VA Lamme. The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of Neuroscience*, 15(2):1605–1615, 1995.

[151] AM Sillito, KL Grieve, HE Jones, J Cudeiro, and J Davis. Visual cortical mechanisms detecting focal orientation discontinuities. *Nature (London)*, 378:492–496, 1995.

[152] HY Lee and M Kardar. Statistics of lines of natural images and implications for visual detection. *arXiv:q-bio.NC/0406051 v1*, 2004.

[153] WJ McGill. Multivariate information transmission. *Psychometrica*, 19(2):97–116, 1954.

[154] RM Fano. *The Transmission of Information: A Statistical Theory of Communication*. MIT Press, Cambridge, Massachussets, 1961.

[155] TS Han. Multiple mutual informations and multiple interactions in frequency data. *Information and Control*, 46(1):26–45, 1980.

[156] N Brenner, SP Strong, R Koberle, W Bialek, and RR van Steveninck. Synergy in a neural code. *Neural Computation*, 12(7):1531–1552, 2000.

[157] RW Yeung. A new outlook on Shannon's information measures. *IEEE Transactions on Information Theory*, 37:466–474, 1991.

[158] D MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[159] S Watanabe. Information theoretical analysis of multivariate correlation. *IBM Journal of Research and Development*, 4:66–82, 1960.

[160] G Miller. *Note on the bias of information estimates II-B*. Information Theory in Psychology. ed. H Quastler (Glencoe, IL: Free Press), 1955.

[161] P Grassberger. Entropy estimates from insufficient samplings. *ArXiv: physics/0307138*, July 2003.

[162] B Efron and RJ Tibshirani. *An Introduction to the Bootstrap*, volume 57 of *Monographs on Statistics and Applied Probability*. CRC Press, London, UK, 1993.

[163] L Schwartz. Theorie des distributions. *Hermann, Paris*, I,II, 1950.

[164] JS Hadamard. Sur les problèmes aux derives partielles el leur signification physique. *Bull. Univ. Princeton*, 13:49–62, 1902.

[165] T Iijima. Basic theory on normalization of a pattern. *Bulletin of Electrotechnical Laboratory*, 26:368–388, 1962.

[166] AP Witkin. Scale-space filtering. In *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, volume 2, pages 1019–1022, Karlsruhe, Germany, 1983.

[167] JJ Koenderink and AJ Doorn. The structure of relief. In *Advances in Imaging and Electron Physics*, number 103, pages 65–150, 1998.

[168] RA Young. Orthogonal basis functions for form vision derived from eigenvector analysis. In *ARVO Abstracts*, page 22, Sarasota, FL., 1978. Association for Research in Vision and Ophthalmology.

[169] RA Young. Gaussian derivative model of spatial vision. *Journal of Optics Society of America A*, 2:102, 1985.

[170] RA Young. Simulation of human retinal function with the Gaussian derivative model. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 564–569, Miami, Fl., 1986.

[171] RA Young. The Gaussian derivative model for spatial vision: I. Retinal mechanisms. *Spatial Vision*, 2(4):273–293, 1987.

[172] M Nielsen. Scale-space generators and functionals. In J Sporring, M Nielsen, L Florack, and P Johansen, editors, *Gaussian Scale-Space Theory*, pages 99–114. Kluwer Academic Publishers, 1997.

[173] BM ter Haar Romeny, WJ Niessen, J Wilting, and LMJ Florack. Differential structure of images: Accuracy of representation. In *Proceedings First IEEE International Conference on Image Processing*, volume 13, pages 21–25, Austin, Texas, 1994.

[174] J Blom, BM ter Harr Romeny, A Bel, and JJ Koenderink. Spatial derivatives and the propagation of noise in gaussian scale-space. *Journal of Visual Communication and Image Representation*, 4(1):1–13, 1993.

[175] LD Griffin, M Lillholm, and M Nielsen. Natural image profiles are most likely to be step edges. *Vision Research*, 44:407–421, 2004.

[176] G Matheron. Modèle sèquentiel de partition alèatoire. *Technical Report, CMM*, 1968.

[177] JP Serra. *Image Analysis and Mathematical Morphology*. Academic Press London, 1982.

[178] L Alvarez, Y Gousseau, and JM Morel. The size of objects in natural and artificial images. *Advances in Imaging and Electron Physics, Academic Press*, 111:167–242, 1999.

[179] AB Lee, D Mumford, and J Huang. Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model. *International Journal of Computer Vision*, 41(1/2):35–59, 2001.

[180] WS Geisler, JS Perry, BJ Super, and DP Gallogly. Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, 41:711–724, 2001.

[181] J Marshall, C Burbeck, and D Ariely. Occlusion edge blur: a cue to relative visual depth. *International Journal of Optical Society of America A*, 13:681–688, 1996.

[182] KA Stevens. *Surface perception from local analysis of texture and contour*. Department of Electrical Engineering and Computer Science, Ph.D thesis. Massachusetts Institute of Technology, 1979.

[183] TH Goldsmith. *Evolution of the eye and visual system. Vision and visual dysfunction*, chapter Evolution of visual pigments and colour vision. Macmillan, London, 1991.

[184] G Osterberg. Topography of the layer of rods and cones in the human retina. *Acta Ophthalmologica*, 6:1–103, 1935.

[185] Curcio CA, Sloan KR, Kalina RE, and Hendrickson AE. Human photoreceptor topography. *Journal of Comparative Neurology*, 292:497–523, 1990.

[186] DA Baylor. Photoreceptor signals and vision. *Investigative Ophtalmology and Visual Science*, 28(Suppl.):34–49, 1987.

[187] RL Gregory. *Eye and Brain*. World University Library, 1966.

[188] DeAngelis GC, Ohzawa I, and Freeman RD. Receptive field dynamics in the central visual pathways. *Trends Neuroscience*, 18:451–458, 1995.

[189] Mumford D. On the computational architecture of the neocortex. i. the role of the thalamo-cortical loop. *Biological Cybernetics*, 65:135–145, 1991.

[190] Mumford D. On the computational architecture of the neocortex. ii. the role of the thalamo-cortical loop. *Biological Cybernetics*, 66:241–251, 1992.

[191] R Tootell, M Silverman, E Switkes, and R De Valois. Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science*, 218:902–904, 1982.

[192] S Levay, DH Hubel, and TN Wiesel. The pattern of ocular dominance columns in macaque visual cortex revealed by a reduced silver stain. *Journal of Comparative Neurology*, 159:559–575, 1975.

[193] DH Hubel and TN Wiesel. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 195:215–243, 1968.

[194] SE Palmer. *Vision Science Photons to Phenomenology*. MIT press, 1999.

[195] S Lehar. Gestalt isomorphism and the primacy of the subjective conscious experience: A Gestalt bubble model. *Behavioral and Brain Sciences*, 26(4):375–444, 2003.

[196] Irvin Rock. *Perception*. Scientific America Books Inc., 1984.

[197] B Julesz. Early vision and focal attention. *Reviews of Modern Physics*, 63(3):735–772, 1991.

[198] I Köhler. *Dynamics in psychology*. Liveright, New York, 1940.

[199] DE Rumelhart, P Smolensky, JL McClelland, and GE Hinton. Schemata and sequential thought processes in pdp models. In *Parallel distributed processing: Explorations in the microstructure of cognition*, volume 2, pages 7–57, Cambridge, MA, 1986. MIT Press.

[200] J Hochberg and D Hardy. Brightness and proximity factors in grouping. *Perceptual and Motor Skills*, 10, 1960.

[201] T Tversky, WS Geisler, and JS Perry. Contour grouping: closure effects are explained by good continuation and proximity. *Vision Research -Oxford-*, 44(24), 2004.

[202] SE Palmer and DJ Levitin. Synchrony: A new principle of perceptual grouping. *39th Annual Meeting of the Psychonomic Society, Dallas, TX*, 1998.

[203] J Hochberg and E McAlister. A quantitative approach to figural "goodness". *Journal of Experimental Psychology*, 46:361–364, 1953.

[204] JF Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:769–798, 1986.

List of Publications

Journals

- AJ Nasrallah and LD Griffin, Gradient direction dependencies in natural images, *Spatial Vision*, 20(3), 277-299, 2007.

- AJ Nasrallah and LD Griffin, The mean power spectrum of an ensemble of scale-invariant images is determined by their mean power spectra, *Image Vision and Computing*, 2006, *Submitted.*

Conferences

- AJ Nasrallah and LD Griffin, Multi-local dependencies of gradient directions in natural images, *Perception* 35(ECVP supplement):57, St.Petersburg, 2006. Poster presentation.

- AJ Nasrallah and LD Griffin, Synergy in the multi-local statistics of gradient directions in images, *CVPRW*, New York, p.173, 20061, 2006. Poster presentation.

- AJ Nasrallah and LD Griffin, Dependencies between gradient directions at multiple locations are determined by the power spectra of the image, Vision in Perception and Cognition, Vol. 35(3), 415–428, *AVA*, Bristol, 20051, 2006. Oral presentation.

- AJ Nasrallah and LD Griffin, Gradient direction dependencies, *NISA Workshop on Feature Learning*, Copenhagen, 2006, Oral presentation.

- AJ Nasrallah and LD Griffin, Statistics of gradient directions in natural images, *BMVA-AVA*, London, 2004. Oral presentation.